Report from Dagstuhl Seminar 21411

Machine Learning in Sports

Edited by

Ulf Brefeld[1], Jesse Davis[2], Martin Lames[3], and James J. Little[4]

1 **Universität Lüneburg, DE,** brefeld@leuphana.de
2 **KU Leuven, BE,** jesse.davis@cs.kuleuven.be
3 **TU München, DE,** martin.lames@tum.de
4 **University of British Columbia – Vancouver, CA,** little@cs.ubc.ca

Abstract

Data about sports have long been the subject of research and analysis by sports scientists. The increasing size and availability of these data have also attracted the attention of researchers in machine learning, computer vision and artificial intelligence. However, these communities rarely interact. This seminar aimed to bring together researchers from these areas to spur an interdisciplinary approach to these problems. The seminar was organized around five different themes that were introduced with tutorial and overview style talks about the key concepts to facilitate knowledge exchange among researchers with different backgrounds and approaches to data-based sports research. These were augmented by more in-depth presentations on specific problems or techniques. There was a panel discussion by practitioners on the difficulties and lessons learned about putting analytics into practice. Finally, we came up with a number of conclusions and next steps.

Seminar October 10–15, 2021 – http://www.dagstuhl.de/21411
2012 ACM Subject Classification Computing methodologies → Machine learning; Computing methodologies → Computer vision
Keywords and phrases machine learning, artificial intelligence, sports science, computer vision, explanations, visualization, tactics, health, biomechanics
Digital Object Identifier 10.4230/DagRep.11.9.45

1 Executive Summary

Ulf Brefeld
Jesse Davis
Martin Lames
Jim Little

Sports has become an incredibly data rich field with the advent of data sources such as event data (e.g., time and locations of actions), tracking data (i.e., positional data), and athlete monitoring (e.g., bio-sensors, IMUs, GPS). These data are commonly and widely collected across multiple different sports, both on a professional and recreational level. The advent of such data raises the need to exploit the collected data both from the theoretical (e.g., sports modeling) as well as practical (e.g., training in top level sports) perspective. Problem-solving solutions can only be provided by an interaction between the sports science & informatics (S&I) and the machine learning (ML) communities. Machine learning is emerging as a powerful, new paradigm for sports analytics, as it provides novel approaches to making sense of the collected data. However, the S&I and ML communities are traditionally separate, each with its own agenda. The seminar aims to bring together top researchers and practitioners who are active in these two fields that can contribute to an assessment of their potential synergies.

We structured the seminar along five different themes, each of which was the focus of between half and a full day. Given the diversity of the participants' backgrounds in terms of discipline, each theme began with overview to get everyone on the same page. Then there were more detailed presentations. The five themes were:

Machine learning meets sports The goal of this session was to provide an overview of some of the machine learning techniques (predictive modeling, text mining) and how they can be applied in the sports context. The illustrative applications where ML can play a role included assessing the performances of teams and players, supporting sport broadcast, assessing fans reactions to rule changes and helping reduce the time burden on video analysts.

Sports science meets machine learning The goal was to provide an overview of basic concepts in sports science to inform researchers from machine learning. The basic concepts were the relation between competition, training and athlete's abilities, the structure of performance in different sports, and the demand for support in sports practice. In particular, the structure of team sports as dynamic interaction processes with emergent behavior was explained as this the most frequent application field for machine learning in sports.

Computer vision for sports The session aimed to expose the participants to the practice of gathering information about team sports through analysis of visual information. The session began with an overview of the general practice of computer vision for sports. Three of the presenters are from industry, representing companies with significant presence in the business of providing to analytics producers information on team sports such as basketball, football (soccer), and ice hockey. The fourth presenter, from academia, discussed material on camera planning and analytics and in addition has himself been involved in tech transfer of visual analytics methods from amateur sports. The overall goal, i.e., informing the participants regarding methods and applications of vision, was well met by the lectures of these experienced researchers.

Interdisciplinary view on tactics The session aimed to build a common understanding of tactics and their implementation in predictive/generative models. It is still an open question how to represent overarching long-term strategies in computer models and different ideas were discussed on the example of overview and contributed presentations.

Explaining, interpreting, and visualizing models and data for sports A key challenge is effectively conveying the results of machine learned models to domain experts, which is compounded by the black-box nature of many such models. This session highlighted a variety of techniques for meeting this objective, with illustrative examples arising from practice were shown for a variety of sports such as ice hockey, table tennis and football. This remains an active area of research and variety of lessons learned and ideas for improving the communication between domain-experts and technical-experts were discussed.

2 Table of Contents

3 Overview of Talks

3.1 Quantifying Offensive Actions, Tactics, and Strategies

Gabriel Anzer (Hertha BSC Berlin, DE)

Offensive performances in football have always been of great focus for fans and clubs alike as evidenced by the fact that nearly all Ballon d'Or winners have been forwards or midfielders. With the increase in availability of granular data, evaluating these performances on a deeper level than just goals scored or gut instinct has become possible. The domain of sports analytics has recently emerged, exploring how applying data science techniques or other statistical methods to sports data can improve decision making within sporting organizations. This thesis follows the footsteps of other sports like baseball or basketball where, at first, offensive performances were analyzed. It consists of four studies exploring various levels of offensive performance, ranging from basic actions to team-level strategy. For that, it uses a data set part of larger research program that also explores the automatic detection of tactical patterns. This dataset mainly consists of positional and event data from eight seasons of the German Bundesliga and German Bundesliga 2 between the seasons 2013/2014 and 2020/2021. In total this amounts to 4, 896 matches, with highly accurate player and ball positions for every moment of the match and detailed logs of every action that occurred, thus making it one of the largest football datasets to be analyzed at this level of granularity. In a first step, this thesis shows how the two different data sources can be synchronized. With this synchronized data it is possible to better quantify individual basic actions like shots or passes. For both actions new metrics (Expected Goals and Expected Passes) were developed, that use the contextual information to quantify the chance quality and passing difficulty. Using this improved quantification of individual actions, the subsequent studies evaluate offensive performance on a tactical pattern level (how goals are scored) and on a strategy level (what team formations are particular effective offensively). Besides their usage on the performance side, these metrics have also been adapted from broadcasters to enhance their data story telling: Expected goals and expected passes are shown during every Bundesliga match to a worldwide audience, thus bringing the field of sports analytics to millions of fans.

3.2 Automated Detection of Complex Tactical Patterns in Football

Pascal Bauer (Deutscher Fußball Bund – Frankfurt am Main, DE)

Football tactics is a topic of public interest, where decisions are predominantly made based on gut instincts from domain-experts. Sport science literature often highlights the need for evidence-based research on football tactics, however the limited capabilities in modelling the dynamics of football has prevented researchers from gaining usable insights. Recent technological advances have made high quality football data more available and affordable. Particularly, positional data providing player and ball coordinates at every instance of a match can be combined with event data containing spatiotemporal information on any event taking place on the pitch (e.g. passes, shots, fouls). On the other hand, the application of machine learning methods to domain-specific problems yields a paradigm shift in many industries including sports. The need for more informed decisions as well as automating time consuming processes–accelerated by the availability of data–has motivated many scientific investigations in football analytics. This thesis is part of a research program combining methodologies from sports and data-science to address the following problems: the synchronization of positional and event data, objectively quantifying offensive actions, as well as the detection of tactical patterns. Although various basic insights from the overall research program are integrated, this thesis focuses primarily on the latter one. Specifically, positional and event data are used to apply machine learning techniques to identify eight established tactical patterns in football: namely high–/mid–/low–block defending, build-up/attacking play in the offense, counterpressing and counterattacks during transitions, and patterns when defending corner-kicks, e.g. player–/zonal– or post-marking. For each pattern, we consolidate definitions with football experts and label large amounts of data manually using video recordings. The inter-annotator reliability is used to ensure that each pattern is well-defined. Unsupervised techniques are used for the purpose of exploration, and supervised machine learning methods based on expert-labeled data for the final detection. As an outlook, semi-supervised methods were used to reduce the labeling effort. This thesis proves that the detection of tactical patterns can optimize everyday processes in professional clubs, and leverage the domain of tactical analysis in sport science by gaining unseen insights. Additionally, we add value to the machine learning domain by evaluating recent methods in supervised and semi-supervised machine learning on challenging, real-world problems.

3.3 From pixels to points: Using tracking data to measure performance in professional sports

Luke Bornn (Zelus Analytics, USA)

In this talk I will explore how players perform, both individually and as a team, on a basketball court. By blending advanced spatiotemporal models with geography-inspired mapping tools, we are able to understand player skill far better than either individual tool allows. Using optical tracking data consisting of hundreds of millions of observations, I will demonstrate these ideas by characterizing defensive skill and decision making in NBA players.

3.4 Understanding the women's football game using machine learning techniques

Lotte Bransen (SciSports – Amersfort, NL)

The women's football game has made huge advances in recent years. Despite its increased popularity, there has been less analysis of data arising from the women's game. This talk provides a summary of our initial work on analyzing the technical data being collected during professional women's football matches. Using event data covering a number of seasons from the top women's leagues, this talk presents three analyses. First, we perform an exploratory analysis by computing several technical indicators (e.g., goal scoring rates over the season, conversion rates, shot locations) and then compare and contrast them to the indicators for comparable men's leagues and find several intriguing differences. Second, we assess whether expected goals (xG) models on one gender are applicable to data from a different gender. Third, we present some preliminary analyses about differences in where and how often women perform certain actions compared to men.

3.5 A computational model for quantifying Availability in soccer

Uwe Dick (Sportec Solutions – Ismanning, DE) and Daniel Link (TU Munich, DE)

The paper presents a computational approach to Availability of soccer players. Availability is defined as the probability that a pass reaches the target player without being intercepted by opponents. Clearly, a computational model for this probability grounds on models for ball dynamics, player movements, and technical skills of the pass giver. Our approach aggregates these quantities for all possible passes to the target player to compute a single Availability value. Empirically, our approach outperforms state-of-the-art competitors using data from 58 professional soccer matches. Moreover, our experiments indicate that the model can even outperform soccer coaches in assessing the availability of soccer players from static images.

3.6 AI videography for amateur hockey

James Elder (York University – Toronto, CA)

Research on computer vision for sports tends to focus on broadcast feeds of professional matches. However, computer vision and AI may have even greater application in amateur sport, for reasons of scale and budget. For hockey, there are 1,000 registered amateur players in North America for every player in the NHL, and amateur leagues typically cannot afford to hire a team of professional videographers for every game. These factors motivate the development of low-cost AI videography systems for amateur hockey; in this talk I will consider three research problems that are central to this goal. The first problem is spatial attention. While hockey is played over a large surface area, instantaneous play is highly localized and fans in the stands use their visual attention and oculomotor systems to continuously follow the play with their eyes. Here I will describe our research efforts to use spatial attention to create a comparable viewing experience for those watching from home. The resulting attentive computer vision system dynamically tracks the play in an 8K wide-field video, automatically extracting a dynamic HD zoomed video streamed to fans in real time. The second problem is temporal attention. While hockey games typically comprise 60 minutes of regulation play, stoppages mean that games take 140 minutes to complete. I will describe a temporal attention system that fuses visual and auditory cues to automatically extract the periods of active play from a recorded video, thus allowing a complete game to be enjoyed offline within a shorter timeframe. The third problem is automatic labeling of players according to team, which is essential for computing play statistics. I will describe an unsupervised learning strategy that allows reliable team assignment even for novel teams, enabling our system to be deployed broadly without retraining.

3.7 AI for sports and health

Björn Eskoffier (FAU Erlangen-Nürnberg, DE)

Wearable computing systems play an increasingly important role in recreational and elite sports. They comprise of two parts. First, sensors for physiological (ECG, EMG, ...) and biomechanical (accelerometer, gyroscope, ...) data recording are embedded into clothes and equipment. Second, embedded microprocessors (e.g. in smartphones) are used for monitoring and analysis of the recorded data. Together, these systems can provide real-time information and feedback for scientific studies in real sports situations. In order to implement these systems, several challenges have to be addressed. Our work focuses on four of the most prevalent of these: (1) Integration: sensors and microprocessors have to be embedded unobtrusively and have to record a variety of signals. (2) Communication: sensors and microprocessors have to communicate in body-area-networks in a secure, safe and energy-saving manner. (3) Interpretation: physiological and biomechanical data have to be interpreted using signal processing and machine learning methods. (4) Simulation and modeling: understanding of sensor data is needed to model processes in sports more accurately, simulation methodologies help here to provide basic information to drive those models. Data mining concepts provide tools for analyzing the considerable amount of physiological and biomechanical data that is generated in sports science studies. Especially when using wearable computing systems, the number of participants and variety of measured data is unlimited in general. Traditional statistical analysis methods commonly cannot handle this amount of data easily. Thus, the analysis is often restricted to individual variables rather than multidimensional dependencies and a considerable amount of information is neglected. Moreover, the results are frequently biased by the expectation of the researcher. Here, the objective, data-driven methods from data mining can contribute by offering useful tools for the analysis tasks. These tools have the ability to deal with large data sets, to analyze multiple dimensions simultaneously, to work data-driven rather than hypothesis-driven, and to provide valuable insights into training effects and injury risks.

3.8 Labeling Situations in Soccer

Dennis Fassmeyer (Leuphana University of Lüneburg, DE)

We study the automatic annotation of situations in soccer games. At first sight, this translates nicely into a standard supervised learning problem. However, in a fully supervised setting, predictive accuracies are supposed to correlate positively with the amount of labeled situations: more labeled training data simply promise better performance. Unfortunately, non-trivially annotated situations in soccer games are scarce, expensive and almost always require human experts; a fully supervised approach appears infeasible. Hence, we split the problem into two parts and learn (i) a meaningful feature representation using variational autoencoders on unlabeled data at large scales and (ii) a large-margin classifier acting in this feature space but utilize only a few (manually) annotated examples of the situation of interest. We propose four different architectures of the variational autoencoder and empirically study the detection of corner kicks, crosses and counterattacks. We observe high predictive accuracies above 90% AUC irrespective of the task.

3.9 Live acquisition of tracking data in soccer

Eric Hayman (ChyronHego – Stockholm, SE)

Tracab is a leading provider of tracking data in soccer, estimating the positions of all players, referees and the ball in real time at 25 times per second. In this talk I will describe our optical tracking system; first launched 15 years ago and which has seen continuous enhancements since. Given that the Dagstuhl seminar contains a number of participants from the sports science community, who consume centre-of-mass optical player tracking data in their work, I will place particular emphasis on the strengths and limitations of optical tracking systems. The motivation is to give those participants a richer understanding of the data they receive. I will highlight how modern deep learning techniques are exploited in our system to enhance both accuracy and automation, but will also emphasize the role of traditional computer vision methods. Moreover, I will describe how human intelligence, in the form of system operators, has been used to complement the strengths of computer algorithms, and how operator workload has been reduced as the technology matures. In the second part of the talk I will describe recent extensions of the system which go beyond centre-of-mass estimation. In particular I will describe how modern deep learning methods permit 3D limb tracking, and how this can be exploited to vastly improve the speed and accuracy of semi-automated offside decision systems to enhance the VAR workflow in soccer. Finally, I will discuss how machine learning techniques can provide a richer stream of player data based on body pose. In the future, this new data should prove beneficial for performance analysis.

3.10 Sport analytics: From pixels to useful metrics

Mehrsan Javan (Sportlogiq, Canada)

Sports analytics is about observing, understanding, and describing the game in an intelligent manner. In practice, most of the focus has been on visual perception to take the video data and extract tracking data and game events. However, turning the incomplete data into actionable insights for the clubs has always been a challenge. This talk focuses on the use of broadcast feed for sport analytics, covers the components of a vision system for data acquisition, provides examples of how Sportlogiq captures the data from broadcast videos and turns them into useful insights for the clubs to make better decisions.

3.11 Providing Personalized training insights in cycling for Team Jumbo Visma

Arno Knobbe (University of Leiden, NL)

In order to gauge how successful their training efforts are, elite sports teams would like to estimate often how fit their athletes are. There are a number of standardized tests that achieve just that, but these tests come with a number of downsides, having to do with the intrusive nature of the tests and the impact that a test has on the intended training schedule. For that reasons, coaches avoid doing tests more than, say, once per month. However, ideally you would be able to get an estimate over several key performance indicators about fitness on a daily basis. This talk describes attempts to achieve this in elite road cycling. Over the last years, power meters have become quite standard in cycling. Combined with the easily obtainable heart rate readings, a fairly detailed source of information becomes available about the work performed on the bike, and the response of the body to this exertion. Normally, one would expect a monotonic relationship between the power output and the heart rate, but with varying exertion over time (for example due to hilly terrain), this relationship is not so clear. We describe techniques to model this data more accurately by integration over time. We will demonstrate that using the right physiological model of the heart's response to external factors, the instantaneous heart rate can be modeled quite effectively. It turns out that parameters of this model fluctuate over time (on the scale of days to weeks), as a function of how tired or fit a rider is, and thus can be used as good indicators of fitness. We give examples of how the heart rate model and its (personalized) parameters can be used to answer several questions in the optimal training of riders.

3.12 Text mining and performance analysis

Otto Kolbinger (TU Munich, DE)

As a discipline originating from notational analysis and biomechanics, performance analysis is dominated by research based on metrical data like match-statistics or physiological measures. However, a lot of evaluations of performance are provided in the form of written text, for example, scouting reports or (pre-) match analysis. Innovative text mining techniques enable researchers to derive knowledge from such written texts in an effective manner but have hardly been applied in performance analysis so far. During my talk, I presented some promising applications of text mining in three areas: the evaluation of technological officiating aids, player scouting and predicting match outcome. For each, I intended to demonstrate how deriving information from text sources can contribute to the knowledge base in the respective area. However, the aim of this talk was twofold. Besides showing such promising approaches, I also pointed at the current lack of universally deployed evaluation standards for classifiers and shortcomings in reporting the performed approaches. An issue not just of text mining – but also of machine learning in applied science in general.

3.13 Sports science for machine learners

Martin Lames (TU Munich, DE)

The idea of this session is to support the interdisciplinary character of the seminar that addresses informaticians as well as sports scientists. Concretely it is about informing informaticians on theories and concepts of sport science. The general areas of training, athletes' capabilities and competition are introduced with a specific emphasis on their interactions. Here, the idea is expressed that support for training can be given by identifying the relevant capabilities responsible for performance in competition. The different structures of performance in the different groups of sports are mentioned, emphasizing that in game sports (team or net sports) there are specific opportunities for machine learning support, because there, we need to decode the relation between action plans and behavioral outcome. The general notion of game sports as dynamic interaction processes with emergent behavior is explained and underpinned with empirical results on the negotiation of match intensity, emergence of behavior, and the impact of chance on goal scoring. The differences between theoretical and practical performance analysis (TPA and PPA) are listed with respect to their main distinguishing aspects. TPA looks for lawlike structures of the sports, whereas in PPA the practical support is dominant. In the former, large data bases on positional and event data create an interesting experimenting ground for informatics, while in the later the challenge lies in the demand for integrating data from many sources in a knowledge base. Each field contains opportunities for machine learning applications if the respective problems are perceived and targeted at.

3.14 The next frontier in sports analytics: Collecting and utilizing tracking data from broadcast video

Patrick Lucey (Stats Perform – Chicago, USA)

Sports Analytics can be divided into two eras: i) the data-driven era, and ii) the AI-driven era. Due to the success of making data-driven decisions, the demand for more granular data and better insights has ushered in the AI-driven era, where computer vision systems are used to collect player/ball tracking data and machine learning to utilize these granular data-sources. Although effective, the AI-Driven era is not yet pervasive due to the hardware requirements of currently deployed computer vision systems (i.e., they need to be in-venue with fixed cameras), which limits the amount of tracking data that can be collected going forward but also historically. However, with the improvement of computer vision technology this is changing. In this talk, I will highlight our work in collecting tracking data from thousands of college basketball broadcast videos and how we can utilize the data collected to make predictions of future NBA talent. I will also highlight current challenges we are working on as we scale out our computer vision solutions. Additionally, I will showcase our recent work in soccer where we predict performance of players on other teams and leagues.

3.15 Diagnostics and training data at the IAT

Björn Mäurer (IAT – Leipzig, DE)

The Institute for Applied Training Science in Leipzig is the central research institute for German elite and junior competitive sport. The main tasks of the institute are performance diagnostics, training analysis and training research. In analysis, we pursue different approaches. Among other things, a 3D analysis of videos is useful and important to be able to analyze the athletes' movements precisely. To speed up the analysis (it is a time-consuming work), we have supplemented the current programme "Mess3d" with a pre-learned software. Here, the athletes' joint points are recorded automatically. Only in case of large deviations of the visual axes of the two cameras (skewed straight lines), a manual check by the scientist is necessary. In this way, a time saving of about 2/3 could be achieved. However, this is very dependent on the type of sport. To achieve a good analysis, you need as much data as possible. These are training, performance diagnostics and competition data. We have developed the IDA software to facilitate the exchange of the necessary data between athletes, coaches and scientists. It enables the institute to obtain the most comprehensive data possible from the athletes. But also to provide coaches and athletes with quick, detailed analyses. The entire digital exchange can thus be handled within one software. This also makes daily communication between all participants much easier. Currently there are about 30 instances of the software with about 5 million completely recorded events. The large amount of data, especially from top-level sport, enables a higher-level analysis. For this task, a machine-based approach is interesting. Therefore, we are currently trying to go down this path. We are always open for cooperation with other scientific institutions.

3.16 Computer vision and AI for sports boadcasting

Fabrizio Pece (Vzrt – Zürich, CH)

Vizrt is a leading provider of cutting-edge solutions for the broadcasting industry. Such solutions allow broadcasters to create and share visually compelling stories, including those concerning live televised sporting events. More specifically, Vizrt's sports production solutions enhance the abilities of broadcasters, coaches, and teams with tools to inform audiences – whether they are at home, in the stands, or in locker rooms. Said tools are built upon state of the art computer vision and machine learning algorithms, and are fused together to create powerful, but yet intuitive solutions to enrich sport event storytelling. In this talk, I will first introduce Vizrt sport production tools, illustrating how computer vision and machine learning algorithm can be used to perform interactive game analysis, real-time, augment-reality graphics insertion, as well as data integration and visualization. The second part of the talk will focus on the challenges of adapting research results to industrial settings, showing how the journey from research to production requires a deep domain understanding to solve domain-specific challenges which are otherwise seldomly tackled in academic settings.

3.17 Winning in sports with data and machine learning

Kostas Pelechrinis (University of Pittsburgh, USA)

Basketball is currently second only to baseball when it comes to integrating data in team operations, decision making and game preparation. While data and statistical analysis have been always part of basketball operations, the availability of detailed player tracking data as well as, additional contextual meta-data, have pushed the envelope further. In this talk, I will start with a specific case study facilitated by player tracking data. In particular, I will present an analysis of how the corner 3-point shots are created and what makes them the second most efficient shot type in the NBA. In the second part of the talk, I will present a more general framework that models the movements of players on the court and tracks the expected points to be scored in real-time. This type of model allows us to evaluate micro-actions such as screens, passes, etc.,that traditionally have been hard to evaluate. Finally, I will briefly discuss other applications of machine learning in basketball.

3.18 HI v. AI in athletic development

Martin Rumo (OYM – Cham, Switzerland)

This presentation argues from the perspective of performance optimization in Sports practice. It explores constraints to the use of machine learning or artificial intelligence concepts in sport practice and argues to formalize human understanding of competition and required skills to enhance the effectiveness of data analytics in sport. Sport is a complex and dynamic domain in which there is a clear goal to optimize nevertheless, the factors which lead to optimal performance are not clear cut. Since sport does not allow any explicit element of luck all factors involved in performance optimization can be considered as skill. Some skill sets, like developing physical strength at a fast rate and doing this repeatedly, are easily measurable and there are lab and field tests to determine those skills in athletes. Since those kinds of basic skills are easily measurable, there is plenty of data available but machine learning concepts do not really contribute to new insights here. Some sports like invasion games are more complex and coaches and athletes elaborate shared concepts of the competition that are not easily represented in data. This introduces a form of semantic gap between the concepts and the available data. This is an interesting area for Machine Learning concepts, but unfortunately the more complex those patterns of events representing the shared concepts become, the less standardized data is available. This is an impediment to training robust Machine Learning models. Furthermore, in sport practice decision making processes underlie an explore-exploit reality, where decisions are not necessarily optimized by exploiting available information, but rather they are motivated by exploring new ways and learning from the resulting outcome. Machine Learning has proven to be very efficient in data generation, especially in computer vision. But using machine learning algorithms in the realm of more abstract structure of the game is more difficult and formalizing human understanding of the underlying structure of the competition seems more effective in providing actionable information for sports practitioner. The presentation explores basic approaches to meet these challenges when delivering real value to the practitioner.

It is recommended to use the discussed approach when practitioners and computer scientists are co-creating valuable data products. Special care should be given to user experience and visualization, since communicating the information is as important in practice as generating it.

3.19 Pitch Control

William Spearman (Liverpool FC, GB)

The presentation introduces recently developed concepts for measuring the importance of space in football. Traditional concepts like controlled zones can be aggregated on a team level to show dominated areas (pitch control). Clearly, controlled space is not equally value. As a proxy, the positioning of the defending team may serve as the basis for computing a measure on space quality. Finally, both metrics can be combined to obtain a computational understanding of control in desired areas of the pitch.

3.20 Biomechanically inspired machine learning to improve performance in biomechanics

Benedicte Vanwanseele (KU Leuven, BE)

The aim of our research is to develop insights as well as tools to come to personalized adaptable rehabilitation and training regimens to enable every individual to physically perform to the best of their own ability. We know that the human body is a complex system where mental, cardiorespiratory and musculoskeletal system is loaded during sport and exercise. In this talk we will focus on the musculoskeletal system as we know that training adaptations to this system are slower compared to the cardiorespiratory system and therefore musculoskeletal overuse injuries are a big challenge in training and sports. To improve performance the human system needs to be load enough but not too much so without increasing the risk of injury. As the musculoskeletal load is the product of volume and magnitude but magnitude is a lot more important it is crucial that we develop method that are able to monitor the load magnitude and how this varies during training. In this talk we will discuss how we develop a method based on a single trunk-based accelerometer to monitor musculoskeletal loading during training and as well as changes within a training session. The second problem is that we cannot really share the code either. The first author, who also implemented everything, has left academia in summer. Unfortunately, the code is in no shape to think of a release in its current form. We will certainly help interested colleagues and may also send the code in private communication but a public release is unfortunately not an option.

3.21 Visual Analytics of Sports Data

Yincai Wu (Zhejiang University - Hangzhou, CN)

With the rapid development of sensing technologies and wearable devices, large sports data have been acquired daily. The data usually implies a wide spectrum of information and rich knowledge about sports. Visual analytics, which facilitates analytical reasoning by interactive visual interfaces, has proven its value in solving various problems. In this talk, I will discuss our research experiences in visual analytics of sports data and introduce several recent studies of our group of making sense of sports data through interactive visualization.

3.22 Visual Analysis of Table Tennis Tactics

Hui Zhang (Zhejiang Universitity – Hangzhou, CN)

Table tennis is a skillful sport, thus, techniques and tactics are the core factors for winning matches. Therefore, coaches, players, and researchers have always paid attention to technique innovations and tactic analysis. In recent years, we have conducted a series of studies on table tennis match data collection, analysis, evaluation, mining, simulation, and video augmentation, mainly including: 1) We proposed a data collection framework to facilitate interactive annotation of table tennis match videos with the support of computer vision algorithms; 2) To further analysis and visual explore the tactics of table tennis matches, a novel interactive visualization system was developed, which provides a holistic visualization of an entire match from three main perspectives, namely, time-oriented, statistical, and tactical analyses; 3) Stroke evaluation is critical for coaches to evaluate players' performance in table tennis matches. For this reason, we proposed an automatic stroke evaluation framework. In particular, to integrate analysts' knowledge into the machine learning model, we employed the latest effective framework named abductive learning, showing promising performance. Based on abductive learning, the system combines the state-of-the-art computer vision algorithms with analysts' knowledge to extract and embed stroke features for evaluation; 4) Tactical analysis in table tennis is challenging as the analysts can often be overwhelmed by the large quantity and high dimension of the data, to address these issues, we designed a visual analytics system to allow analysts to effectively analyze, explore, and compare tactics of multiple matches based on the advanced embedding and dimension reduction algorithms along with an interactive glyph; 5) Simulative analysis in competitive sports can provide prospective insights, which can help improve the performance of players in future matches. We propose a well-established hybrid second-order Markov chain model to characterize and simulate the competition process in table tennis. Compared with existing methods, our approach supports the effective simulation of tactics, which represent high-level competition strategies in table tennis; 6) Visualizing data in sports videos is gaining traction in sports analytics, given its ability to communicate insights and explicate player strategies engagingly. So, we design fast prototyping tool, to ease the creation of augmented table tennis videos by leveraging machine learning-based data extractors and design space-based visualization recommendations. With the system, analysts can create an augmented video by selecting the table tennis match data to visualize instead of manually drawing the graphical marks.

3.23 Using machine learning to assess and compare athletes in team sports

Albrecht Zimmermann (University of Caen, FR)

Using machine learning techniques (ML) to assess action (and derived from it player) quality is a recent and promising alternative to expert-based assessments. The big challenges for these approaches consist of data acquisition, data transformation and augmentation, which often require in-depth knowledge of the sport in question, as well as understanding which ML techniques are well-suited not only for final modeling but also for data preparation.

4 Panel discussions

We organized a panel on the challenges of putting data science insights into practice. Therefore, we asked a number of attendees with experience in this regard to serve as panelists: Luke Bornn (Zelus Analytics), Lotte Bransen (SciSports), Jan Van Haaren (FC Brugge), Mehrsan Javan (Sportlogic), Patrick Lucey (STATS), Benedicte Vanwanseele (KU Leuven).

Before the panel, we solicited topics from the attendees. Based on the response we posed the following questions to the panel:

- What are the key challenges and opportunities for applying machine learning to sports problems in sports beyond football?
- It is often argued that each athlete (or team) is unique, which indicates that a personalized modeling approach would be ideal. However, machine learning works best when there is lots of data, which suggests pooling data across athletes is the way to go. How do you navigate this trade-off between developing a personalized model vs. one model that serves all?
- What challenges did you face when communicating insights from data to practitioners? How did you overcome them?
- What are the key open opportunities for machine learning in physiological monitoring?
- What are the key open opportunities for machine learning in tactical analyses?
- Who do you see as key end users of the advanced analyses? For example fans, teams, coaches, media, ...?
- Particularly for physical monitoring, we have advanced techniques that yield very accurate measurements in a lab setting. However, measuring in the wild during competitions is more challenging. How do we know that what that we find in the lab translates to a more ecological setting?

5 Results

During the seminar, we identified the following actions and activities to establish a forum for future exchange and bringing the disciplines together:

1. We endeavor to participate in each other's regular conferences and meeting such as International Symposium on Computer Science in Sport, Machine Learning and Data Mining for Sports Analytics, Computer Vision for Sports, etc. We will try to facilitate combined events to continue the interaction.

2. To facilitate this initiative, we will set up a mailing list of those that are interested in conferences, seminars, job opportunities, etc. The list is: ml-ai-4sports@googlegroups.com
3. We will aim to setup a reoccurring online seminar series with sessions every four to six weeks.
4. We will explore the possibility of securing funding to establish a network. Arnold Baca has some experience in this regard and graciously offered to investigate it.

Participants

- Gabriel Anzer
Hertha BSC – Berlin, DE
- Arnold Baca
Universität Wien, AT
- Pascal Bauer
DFB – Frankfurt, DE
- Ulf Brefeld
Universität Lüneburg, DE
- Jesse Davis
KU Leuven, BE
- Björn Eskofier
Universität Erlangen-Nürnberg, DE
- Dennis Faßmeyer
Leuphana Universität Lüneburg, DE
- Eric Hayman
ChyronHego – Stockholm, SE
- Arno J. Knobbe
Leiden University, NL
- Otto Kolbinger
TU München, DE
- Philipp Kornfeind
Universität Wien, AT
- Martin Lames
TU München, DE
- Daniel Link
TU München, DE
- Björn Mäurer
IAT – Leipzig, DE
- Fabrizio Pece
Vizrt AG – Zürich, CH
- Martin Rumo
OYM AG – Cham, CH
- Tiago Guedes Russomanno
University of Brasilia, BR
- Marc Schmid
TU München, DE
- Jan Van Haaren
FC Bruges, BE
- Benedicte Vanwanseele
KU Leuven, BE
- Albrecht Zimmermann
Caen University, FR

Remote Participants

- Luke Bornn
Simon Fraser University – Burnaby, CA
- Lotte Bransen
SciSports – Amersfoort, NL
- Mirjam Bruinsma
AFC Ajax – Amsterdam, NL
- David Carey
La Trobe University – Melbourne, AU
- Xiangtong Chu
Zhejiang University, CN
- Laura de Jong
Deakin University – Melbourne, AU
- Uwe Dick
Sportec Solutions AG – Ismaning, DE
- James Elder
York University – Toronto, CA
- Irfan A. Essa
Georgia Institute of Technology – Atlanta, US
- Mehrsan Javan
Sportlogiq – Montréal, CA
- Jim Little
University of British Columbia – Vancouver, CA
- Patrick Lucey
STATS Perform – Chicago, US
- Konstantinos Pelechrinis
University of Pittsburgh, US
- Eraldo Luis Rezende Fernandes
Leuphana Universität Lüneburg, DE

- Karen Roemer
Central Washington University – Ellensburg, US
- Yannick Rudolph
Leuphana Universität Lüneburg, DE & SAP SE – Berlin, DE
- Oliver Schulte
Simon Fraser University – Burnaby, CA
- William Spearman
Liverpool Football Club, GB
- Karl Tuyls
DeepMind – London, GB
- Maaike Van Roy
KU Leuven, BE
- Jiang Wu
Zhejiang University – Hangzhou, CN
- Yingcai Wu
Zhejiang University – Hangzhou, CN
- Hui Zhang
Zhejiang University – Hangzhou, CN
- Zhou Zheng
Zhejiang University – Hangzhou, CN