Computer Science in Sport – Research and Practice: A book review

Stuart Morgan

Movement Science, Australian Institute of Sport

*Department of Computer Science and Information Technology, La Trobe University,
Australia*

Abstract

Sports informatics and computer science in sport are perhaps the most exciting and fast-moving disciplines across all of sports science. The tremendous parallel growth in digital technology, non-invasive sensor devices, computer vision and machine learning have empowered sports analytics in ways perhaps never seen before. This growth provides great challenges for new entrants and seasoned veterans of sports analytics alike. Keeping pace with new technological innovations requires a thorough and systematic understanding of many diverse topics from computer programming, to database design, machine learning algorithms and sensor technology. Nevertheless, as quickly as the state of the art technology changes, the foundation skills and knowledge about computer science in sport are lasting. Furthermore, resources for students and practitioners across this range of areas are scarce, and the new-release textbook *Computer Science in Sport: Research and Practice* edited by Professor Arnold Baca, provides much of the foundation knowledge required for working in sports informatics. This is certainly a comprehensive text that will be a valuable resource for many readers.

KEYWORDS: COMPUTER SCIENCE, INFORMATICS, REVIEW, ANALYTICS

Introduction

In modern science, no field of endeavor is ever truly independent of other fields, and this is especially true of the rapidly evolving field of informatics and computer science in sport. Notational analysis, the historical study of patterns, tactics and strategy, is unrecognizable from its early form and dependence on manual (pencil and paper) data collection and summary statistics. The evolution of technical tools, primarily in the domain of computers and digital video, have led to profound changes in notational analysis and performance analysis, and

ultimately to the birth of arguably a new field of sports science in its own right – sports informatics. In this review paper I discuss the content of a recently released textbook for students and practitioners in computer science in sport, edited by Professor Arnold Baca (Baca, 2015). I will argue that modern sports informatics is enabled by dramatic parallel developments in computer science, machine learning and even artificial intelligence, and that the textbook addresses much of the foundation knowledge required for students and researchers to engage actively in this field of enquiry.

One may argue about the semantics in the naming of a research field, but I will use computer science in sport interchangeably with sports informatics. Even though in practice research in computer science in sport may address areas that are not strictly sports informatics (such as computer vision or machine learning), and sports informatics research reaches to areas less dependent on computer science, there is sufficient overlap in the context of this book review to consider both as being at the vanguard of exciting new research that enables all of the traditional fields of sports science.

Link and Lames (Chapter 1: *An Introduction to Sports Informatics*) present the opening chapter of the book, with an historical context of the field of notational analysis from the mid 1970's. This history also describes the origin of the International Association for Computer Science in Sport (IACSS) in 2002, which has been largely the only formal organization supporting a community of practice in sports informatics since that time. Link and Lames emphasize the interdisciplinary nature of sports informatics, highlighting also the attraction of computer scientists to sport due to the “...existing structures [that are] are neither too simple to be of interest, nor too complex to be described using mathematical models” (p7), which is an essential point. Sport is dynamic and uncertain, yet there are undoubtable regularities in athlete behavior and performance. The availability of large-scale, highly-dimensional sports data presents many challenges to researchers at the very edge of contemporary computer science to reveal insights about complex cooperative and competitive behavior by athletes. Recent and impactful papers include latent factor models and fine-grained prediction in basketball (Yue *et al*, 2015), and deep learning, where convolutional neural networks (Karparthy *et al*, 2014) and neural networks with short-term memory (Yue-Hei Ng *et al*, 2015), are used to classify visual features in fast-moving sports.

Following this introduction Dabincki & Miyaji (Chapter 2: *Computers, Informatics and Sport*) offer a foundation and historical perspective on computer architecture and database systems. Lyons (Chapter 3: *Databases and Expert Systems*) also describes some of the concepts prevalent in early IF-THEN “expert” systems. Lyons also draws on his own experiences in a national sports institute to reflect on the nexus between computer science expertise and the realm of the sports coach, and proposes an imaginary “institute” where domain experts in sports and informatics work seamlessly together. More frequently, however, it is coaches and sports scientists with minimal computer science training who are responsible for the design and implementation of database systems in sport. In that light, this chapter provides useful framework for database design, and a backdrop to more sophisticated modern algorithms that now pervade the field of machine learning.

If sophisticated algorithms are the engine for gaining insight in sports informatics, then data is the fuel. High quality data is at the core of any meaningful attempt to understand behavior and causality in sports performance. The next chapter by Baca (Chapter 4: *Data Acquisition and Processing*) discusses a vast range of sensor types capable of producing data in the sports domain (such as accelerometers, GPS, heart rate monitors, and other biometric monitors). Each sensor type has its own signal-response characteristics and it is critical for any researcher or practitioner to understand those characteristics when interpreting their data. This is an

excellent chapter, which offers an oft-neglected discussion about “how good is my data?”, and “what should I know about my data?” to avoid errors of interpretation such as aliasing or under sampling.

Moreover, modern sports informatics is characterized by two additional challenges; to measure performance *in-situ* without impacting on the performance itself, and enabling real-time analysis of performance in such a way that a coach (or athlete) is able to moderate and optimize an ongoing performance. Baca refers to the technology underpinning these *in-situ* tools and real-time feedback as “ubiquitous computing”. In both challenges, sensors and data transmission are fundamental concepts to understand. Baca also addresses data transmission protocols including USB, Firewire 1394, LAN, Bluetooth, ANT and ZigBee. Again, each of these protocols present their own transmission characteristics and hardware requirements, and a good sports informatics specialist should have a sound understanding of the technical tools available in their domain.

Signal processing is a well-understood area for engineers, however, the training for sports scientists often overlooks topics that are vital for a sports informatics specialist. Data normalization, and temporal resampling (for instance) are key concepts in basic signal processing, but additional techniques such as Fast Fourier Transform (FFT), which allows data to translate between the spatial and temporal domains are highly valuable tools in the sports analysts toolkit. The Baca chapter covers essential material for a basic understanding of these data treatments, when and why they should be used, and some pitfalls for the unwary on how to accurately interpret data streams.

Leser & Roemer (Chapter 5: *Motion Tracking and Analysis Systems*) rightly state that player tracking systems have been the most profound technical innovation for team sports in recent decades. However, the technology provides more data than coaches (in many cases) can use. For this reason, one of the main goals for sports informatics as a discipline, and certainly for the performance analyst in the field, is to find actionable intelligence from great volumes of data, and to distill insight from large-scale data without defaulting to summary statistics or the mean value. One approach to this problem is to visualize data in creative ways to allow coaches to simplify their search, allowing them to focus on a smaller number of specific events, and then to use their domain expertise at a qualitative level to gather insight. Leser & Roemer cover some of the data visualization techniques that allow coaches to precisely that.

The science of sports informatics, however, demands more than to simply visualize data and redirect a coach to frequently occurring events. Perl (Chapter 6: *Modeling and Simulation*) has pioneered neural networks in sports (e.g. Memmert & Perl, 2009, Silva *et al*, 2007), and his chapter on modeling and simulation covers a great deal of essential knowledge in this exciting area. The essential properties of a model are data compression (dimensionality reduction), and prediction. Arguably, a model is only useful if it can do both. This is a complex area, and readers should take their time to explore this detailed chapter. Perl addresses foundation principals and worked examples in dynamic systems, Monte Carlo simulations, and touches broadly on artificial intelligence.

Oddly perhaps, some of the most exciting and rapidly evolving areas in computer science have been grouped as “*unconventional modeling*” (p 137). It is a contestable definition. In recent months and years deep learning approaches to artificial intelligence have profoundly changed the way computer scientists think about neural networks. Some early approaches such as evolutionary algorithms and simple neural networks have arguably become outdated, outmuscled by deep learning algorithms. That some of these early methods are discussed in this chapter is less of a criticism of Perl, and more of a point to highlight the extraordinary rate

of change in this domain.

O'Donoghue (Chapter 7: *Game Analysis*) discusses game analysis systems including a general design architecture for creating bespoke applications. In practice, coaches and analysts must often choose between general-purpose systems that serve broadly as video-indexing tools, and customized match analysis systems designed for more specific purposes. The latter may be developed fit for an explicit purpose in a single sport where the format of the sport requires particular analysis features. Although such systems can be quite comprehensive, it is debatable that they could be regarded as new contributions to computer science. It is often desirable for an aspiring sports analyst that he/she is capable of sufficient programming in a high level language such as Visual Basic or Python, to compile a useful game analysis system (O'Donoghue's chapter provides sound guidelines for doing that), however, the frontier of computer science in sport is far more advanced to exploit modern computing power, machine learning algorithms, and state of the art computer vision. At the vanguard, the challenges in computer science are to find meaningful insight and sports-specific decision support from large-scale spatio-temporal data, highly dimensioned data sources, and at an even more fundamental level, unstructured pixel data. Significant recent work has progressed the field of computer vision (where sport is a popular and challenging context for research), and the analysis of the resultant spatio-temporal data, but unfortunately most of that literature is absent in this book. There is, however, a comprehensive guide to elementary computing and system design that many students of sports informatics should find very helpful.

Much of the sports informatics domain is about the interface between coaches, data, and maximizing the insight that the former can glean from the latter. But another area that is less-frequently considered by sport-focused computer scientists and data specialists is that of optimizing the learning and training of the athlete. The final chapter by Wiemeyer and Mueller (Chapter 8) covers a range of topics under the heading of *information and communication technology-enhanced learning and training* (ICT-ELT). This is an intriguing concept, since it implies that learning can be improved (or enhanced) by appropriate ICT-mediated approaches. In fact, Wiemeyer and Mueller propose explicitly that ICT-ELT should not only match the efficacy of "traditional" learning, but should also enhance learning to a degree that justifies the technological resources used. This is a very detailed chapter that explores various modalities of ICT-ELT, such as augmented reality, virtual reality, video technologies, and mobile measurement systems (not an exhaustive list). These are all very important modes of information delivery and user-engagement/interaction, and the chapter discusses the characteristics of the technology (i.e. real time, *in-situ*, simulated, interactive), and important concepts about motor learning, skill acquisition and perceptual expertise, which are critical concepts to understand for appropriate use. This chapter provides a very neat domain on which to round out the coverage of the broader text.

Conclusion

Computer Science in Sport: Research and Practice is certainly a comprehensive collection of foundation principles relevant to all sports informatics practitioners and students. It also provides an excellent resource for seasoned researchers in sports science who want to delve into the exciting and fast moving sports analytics domain. Returning to the observation by Lames and Link in the opening chapter, sport represents a domain that is complex enough to be interesting to computer and data science researchers, but it is by no means simple. Indeed, powerful recent discoveries in artificial intelligence are only just starting scratch the surface of the rich array of insight possible in modern sports. Nevertheless, even the most complex mathematical models must still be predicated on sound basic science, and each chapter

contributes to a thorough framework of understanding woven from a mix of research and practice. This book should enable readers to take their first confident steps into the exciting world of sports informatics and the computer sciences.

References

- Baca, A. (2015). *Computer Science in Sport: Research and Practice*. London and New York: Routledge.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). *Large-scale Video Classification with Convolutional Neural Networks*. In Proceedings of International Computer Vision and Pattern Recognition (CVPR 2014).
- Memmert, D., & Perl, J. (2009). Game creativity analysis using neural networks. *Journal of Sports Sciences*, 27(2), 139-149.
- Silva, A. J., Costa, A. M., Oliveira, P. M., Reis, V. M., Saavedra, J., Perl, J., & Marinho, D. A. (2007). The use of neural network technology to model swimming performance. *Journal of Sports Science and Medicine*, 6(1), 117-125.
- Yue, Y., Lucey, P., Carr, P., Bialkowski, A., & Matthews, I. (2014). *Learning Fine-Grained Spatial Models for Dynamic Sports Play Prediction*. In 2014 IEEE International Conference on Data Mining, Shenzhen, pp. 670-679.
- Yue-Hei Ng, J., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., & Toderici, G. (2015). *Beyond Short Snippets: Deep Networks for Video Classification*. In Proceedings of International Computer Vision and Pattern Recognition (CVPR 2015). pp. 4694-4702.