

MATHSPORT INTERNATIONAL 2019 CONFERENCE

-Book of Astracts -

Crowne Plaza Hotel Athens, 1-3 July 2019



Organized by AUEB Sports Analytics Group

Conference Sponsors







Original title Proceedings of MathSport International 2019 Conference Editors: Dimitris Karlis, Ioannis Ntzoufras, Sotiris Drikos 2019 Athens University of Economics and Business AUEB Sports Analytics Group AUEB Computational and Bayesian Statistics Lab.

BOOK OF ABSTRACTS



7TH MATHSPORTS INTERNATIONAL CONFERENCE

ATHENS, 1-3 JULY 2019

Preface

We are very glad to welcome you at Athens for the 7th Mathematics in Sports International conference. It is our pleasure to host the conference this year after the successful previous ones in Manchseter, Groningen, Manchester, Leuven, Loughborough and Padova. Since the first MathsSports conference a lot of research has been made on topics related to Mathematics and Sports and the interest has increased considerably. MathsSports conference is a gathering of Academics and practitioners covering a wide range of sports and methodologies and we hope that this year the meeting will cover even more sports. We have attempted to arrange a wide range of talks during the 3 days of the conference.

A complete list of all the abstracts of the papers to be presented in the conference can be found in this book. A detailed index of all authors can be found at the end to facilitate easy search.

We hope that you will enjoy the meeting

Dimitris Karlis Ioannis Ntzoufras on behalf of the LOC and SC.

Scientific Committee

Phil Scarf	University of Salford, Manchester (UK)
Anthony Bedford	RMIT University, Melbourne (AU)
Marco Ferrante	University of Padova (IT)
Dries Goossens	Gent University (BE)
Ruud Koning	University of Groningen (NL)
Stephanie Kovalchik	Victoria University (AU)
Alun Owen	De Montfort University - DMU Leicester (UK)
Frits Spieksma	Eindhoven University of Technology (NL)
Ray Stefani	CSULB University (USA)

Local Organizing Committee

Dimitris Karlis Ioannis Ntzoufras	Chair Athens University of Economics and Business Athens University of Economics and Business
	Members
Sotiris Drikos	AUEB Sports Analytics Group (GR)
Panagiotis Repousis	Athens University of Economics and Business (GR)
Evgenia Tzoumaka	American College of Greece (GR)
Konstantinos Matzorakis	AUEB Sports Analytics Group (GR)

Contents

The influence of match phase and field position on collective team behaviour in Australian Rules football Jeremy Alexander, Bartholomew Spencer, Alice Sweeting, Jocelyn Mara, Sam	
Robertson	1
Rest Differences among Teams in European Football Leagues Tankut Atan, Burak Cavdaroglu	2
Alternative count regression models for modeling football outcomes Alessandro Barbiero	3
Econometric approach to assess the transfer value of professional football players	4
Roger Besson, Raffaele Poli, Loic Ravenel	4
Coaching Athletes with Hidden Disabilities: Collaborative research findings and measurement <i>Robbi Beyer</i>	5
Using a multi-camera tracking system to estimate ball spin in tennis Olivia Cant, Stephanie Kovalchik, Rod Cross, Machar Reid	6
Integrated Break and Carryover Minimization Problem in Round Robin Tournaments	7
Burak Cavaarogiu, Tankut Atan	(
Modelling the outcomes of professional snooker matches James Collingwood, Michael Wright, Roger Brooks	8
Overcoming the incentive incompatibility of tournaments with multiple	
group stages Laszlo Csato	9
Comparison between some European football leagues through related count data variables	
Valentina Cueva, José Rodríguez-Avi, María José Olmo-Jiménez	10
Randomness of Play Calling in College Football Brian Curley, Gretchen Hopkirk, Ryan Lokhorst, Annette Pilkington	11
A Modern Love Story: Machine Learning and the Global Sports Betting Industry	
Lloyd Danzig	12

Establishing a performance edge in P2P betting Carl Donovan, Bruno Caneco, Colin Bleak	13
Complex 1 in Male Volleyball as a Markov Chain Sotiris Drikos	14
Modelling volleyball data using a Bayesian approach Leonardo Egidi, Ioannis Ntzoufras	15
Extreme value prediction: an application to sport records Giovanni Fonseca, Federica Giummolé	16
Player impact measures for scoring in ice hockey Carles Sans Fuentes, Niklas Carlsson, Patrick Lambrix	17
Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data <i>Floris Goes, Matthias Kempe, Koen Lemmink</i>	18
Champions League or domestic league: a coach's choice Dries Goosens, Chang Wang, Martina Vandebroek	19
Play-by-play data analysis for team managing in basketball Luca Grassetti, Ruggero Bellio, Giovanni Fonseca, Paolo Vidoni	20
Fair-Fixture: Minimizing Carry-Over Effects In Football Leagues Dilek Günneç and Ezgi Demir	21
Will Groups of 3 Ruin the World Cup? Julien Guyon	22
Analysing the effect of a change of transition probabilities related to possession on scoring a goal in a football match Nobuyoshi Hirotsu, Ayako Komine	23
Score-based soccer match outcome modelling - an experimental review Ondřej Hubáček, Gustav Šourek, Filip Železný	24
Modelling the 2018 FIFA World Cup - Could we have "Beaten the Bookies" ?	
Gordon Hunter, Benjamin Jauvion, Ishan Rashid, Mohamed Sharif-Ali	25
Using assessments to facilitate mental performance consulting Madison Hunt, Anastasia Blevins	26
Serve and Return Glicko: An approximate Bayesian model for within-match win prediction Martin Ingram	27
Sports Coaching: An Art and a Science Simon Jenkins	28
The explosive growth of eSports and the potential for research opportunities	
Lyn Kee, Minh Huynh, Denny Meyer, Kelly Marshall	29
Move it or lose it: Exploring the relation of defensive disruptiveness and team success Matthias Kempe, Floris Goes	30

An evaluation of the three-point rule in football Wim Koevoets	31
A quantitative method for evaluating the skills of national volleyball teams: Prediction accuracy comparisons of the official ranking system in the worldwide tournaments of 2010s	
Eiji Konaka	32
Availability-to-train in elite sport Ioannis Kosmidis	33
Random Walks with Memory Applied to Grand Slam Tennis Matches Modeling Tomáš Kouřim	34
Extensions of the Elo Rating System for Margin of Victory Stephanie Kovalchik	35
Fair sudden death Roel Lambers, Frits Spieksma	36
On the flexibility of Home-Away pattern sets Roel Lambers, Dries Goossens, Frits Spieksma	37
The Age Advantage in Youth Football Steve Lawrence, Laura Jonker, Jan Verbeek	38
Survival Modelling of Goal Arrival Times in Champions League Ilias Leriou, Ioannis Ntzoufras, Dimitris Karlis	39
Talent detection in sport : Machine Learning methods for performance prediction Arthur Leroy	40
Favorite longshot bias and betting margin: An insider's perspective Kostas Maras	41
Assessing Mental Toughness Leilani Madrigal	42
Maximizing financial and on-field performances when composing teams in soccer	
Lars Magnus Hvattum, Giovanni Pantuso	43
Passing tactical relevance during soccer matches determined by machine learning Murila Marlin, Falina Maura, Piacada Tarma, Vitan Princing, Samia Augusta Cumba	4.4
Murilo Merin, Felipe Moura, Ricardo Torres, Vilor Frincipe, Sergio Augusto Canna	44
Fame and Fortune in Elite Tennis Revisited Denny Meyer, Minh Huynh, Kelly Marshall, Geoff Pollard	45
Flexible multivariate point processes with applications to modelling football matches	40
Sanmosn Ivarayanan, Ioannis Kosmiais, Petros Dellaportas	40
Lessons Learned in Scheduling the Finnish Major Ice Hockey League Kimmo Nurmi, Jari Kyngäs, Nico Kyngäs	47

New Performance Measures for the NBA: An Indirect Evolutionary Game Theory Approach John Obrien, Mark Fichman	48
Modelling Momentum in Football Marius Ötting, Roland Langrock, Antonello Maruotti	49
Statistical Models of Horse Racing Outcomes Using R Alun Owen	50
Bayesian Modelling of Volleyball Sets Vasileios Palaskas, Ioannis Ntzoufras, Sotiris Drikos	51
Quantifying the Evolution of First-Class Rugby in New Zealand Ankit Patel, Paul Bracewell, Jack McIvor	52
Detection of Playing Style of Football Teams and Players using Latent Dirichlet Allocation David Perdomo Meza	53
Winning in Basketball with Data and Machine Learning Konstantinos Pelechrinis	54
Multi-Criteria Solutions for Optimizing Lineup in Baseball Marina Polyashuk	55
Perspectives on Fairness in Sport Chris Potts	56
Finding optimal strategies for substitutions in soccer using a two-scale Markov Decision Process approach Jörg Rambau, Rónán Richter	57
Anti-Analytics in Football Anay Rennie, Jayanth Kolla	58
Iterated Local Search for Multi-league Scheduling Problems Panagiotis Repoussis, Walter Gisler	59
A Bayesian dynamical bivariate Poisson state space model for predicting football scores and results	
Peter Ridall	60
Modelling netball scores Phil Scarf, Rose Baker, Rishikesh Parma	61
Predicting possession outcomes using spatiotemporal data in Australian Rules football Bartholomew Spencer, Karl Jackson, Sam Robertson	62
Barwoolonical Spencer, Harry season, Sam Robertson Ranking Elite Swimmers using Extreme Value Theory Harry Spearing, Jonathan Tawn	63
The Ancient Olympics: Events, Technology, Superstars, Women, Lessons for Them and for Us <i>Raymond Stefani</i> .	64
Raking of Japanese University Baseball Takashi Toriumi	65

Objectively Modelling the College Football Playoff Committee's Selections John Trono	66
Combining the Four Factors with the Generalized PageRank (GeM) model for NBA Basketball Georgia Twersky, George Lyman, Anne Pilkington	67
Using the Poisson-Lognormal Regression Model with Varying Dispersion to analyze soccer data George Tzougas, Dimitris Karlis	68
Sports betting strategies: an experimental review Matej Uhrín, Gustav Šourek, Ondřej Hubáček, Filip Železný	69
Identifying Technical Functions of Footballers using Hierarchical Cluster Analysis Orbay Unsoy	70
A hybrid random forest to predict soccer matches in international tournaments Hans Van Eetvelde, Christophe Ley, Andreas Groll, Gunther Schauberger	71
Pre-Game Speeches: Impact and Measurement Tiffanye Vargas	72
Avoiding combinatorial clashes for the Champions Hockey League Group Stage Draw Stephan Westphal, Martin Dahmen	73
Contextual information improves the accuracy of predicting the direction of serve in professional tennis Hiroyuki Yamamoto, Kazutoshi Kudo, Tim Buszard, Machar Reid, Damian Farrow, Stephanie Kovalchik	74
Structuring Patterns in Team Movement and Performance from Vector Field Analysis and Fractal Dynamics Abdullah Zafar, Farzad Yousefian	75
	76

Abstracts

The influence of match phase and field position on collective team behaviour in Australian Rules football

Jeremy Alexander¹, Bartholomew Spencer¹, Alice Sweeting¹, Jocelyn Mara², and Sam Robertson¹

¹ Victoria University, Australia

² University of Canberra, Australia jeremyalexander60@gmail.com b.spencer89@gmail.com jocelyn.mara@canberra.edu.au Sam.Robertson@vu.edu.au

Abstract. Introduction Collective team behaviour in invasion sports has been used to describe team tactics, whereby repetitive patterns of movement are formed (1). This behaviour has typically been defined via various spatiotemporal metrics including x-axis centroid, y-axis centroid, length, width, and surface area (2). More recently, studies have utilised heat maps combined with entropy measures to determine and visualise the variability of player movement. The resulting information has become a central component of match analysis due to its established relationship with performance outcomes and the capability to provide greater context to match events (2).

Methods Data from professional male athletes (years 24.4 ± 3.7 ; cm 185.9 ± 7.1 ; kg 85.4 ± 7.1), were collected via 10 Hz global positioning system (GPS) during a competitive AFL match. Five spatiotemporal metrics (x-axis centroid, y'axis centroid, length, width, and surface area), heat maps, and Shannon Entropy were analysed by match phase (offensive, defensive, and contested) and field position (defensive 50, defensive midfield, forward midfield, and forward 50). A multivariate analysis of variance (MANOVA) was assessed to compare between match phase and field position. Variability of player movement was quantified via heat maps and Shannon Entropy.

Results Field position had a greater influence on the x-axis centroid comparative to match phase. Conversely, match phase had a greater influence on length, width, and surface area when compared with field position. Heat maps displayed that players repositioned behind centre when the ball was closer to their goal and pushed higher up the field when the ball was in their forward half. Shannon Entropy showed player movement was more variable during offence and defence compared to contested phases.

Conclusion Analysis of spatiotemporal metrics, heat maps, and Shannon Entropy whilst accounting for contextual variables may provide a greater understanding of the game style of AF teams.

Keywords

Performance Analysis, Game Style, Team Tactics

Rest Differences among Teams in European Football Leagues

Tankut Atan¹ and Burak Cavdaroglu²

- ¹ Bahcesehir University, Turkey
- ² Kadir Has University, Turkey sabritankut.atan@eng.bau.edu.tr burak.cavdaroglu@khas.edu.tr http://www.khas.edu.tr/cv/2685

Abstract. When two opponents in a sports game have not rested an equal amount after their most recent game, the more rested team has an unfair advantage over the less rested team. Tournament organizers typically do not pay attention to this fairness criterion when they determine the timings of the games. We provide two general linear integer programming formulations for a given round robin tournament schedule that find the periods of the games minimizing the total rest difference and the number of occurrences of rest difference among the teams. Then, we compare how the top division professional football leagues in Europe perform in terms of these objectives.

Keywords

Integer Programming, Timetabling, Tournament Fairness

Alternative count regression models for modeling football outcomes

Alessandro Barbiero

Department of Economics, Management and Quantitative Methods, Universitá degli Studi di Milano, Italy alessandro.barbiero@unimi.it

Abstract. In this work, we propose the use of discrete counterparts of the Weibull distribution for modeling football results, as an alternative to existing bivariate Poisson regression models and modifications/extensions thereof, such as diagonally inflated or generalized Poisson models. The simple bivariate Poisson model, with independent components, was the first used in football data analysis for modelling the outcome of a game (number of goals scored by the two competing teams) due to its ease of use and interpretation. Later, more complex models allowing for non-null correlation were explored, since real data often show a slight but non-negligible positive correlation between the numbers of goals scored by the two teams; or allowing for overdispersion and excess in draws, which usually characterize football outcomes. The discrete Weibull distributions derived as analogues of the homonym continuous distribution seem to be more flexible than Poisson, since adjusting their two parameters can model a variety of different features. The numbers of goals scored by the two teams can be regarded as a joint observation from a bivariate random vector with discrete Weibull margins, linked through a copula function that accommodates dependence. The parameters of the distribution are assumed to depend on covariates such as the attack and defense abilities of the two teams and the "home effect". Several discrete Weibull regression models are proposed, by varying the type of discretization, the copula function, the choice of covariates, and are then applied to the Italian Serie A championship. Even if the interpretation of parameters is less immediate than in Poisson models, yet they represent a suitable alternative, as the application demonstrates, and can be employed as a statistical tool for better understanding the performance of teams in order to improve predictions, from a betting perspective, or to deploy corrective actions, from a managerial point of view.

Keywords

count regression model, discrete Weibull distribution, statistical dependence

Econometric approach to assess the transfer value of professional football players

Roger Besson, Raffaele Poli, and Loïc Ravenel

CIES, Switzerland roger.besson@unine.ch raffaele.poli@unine.ch loic.ravenel@unine.ch

Abstract. This communication outlines the approach developed by the CIES Football Observatory research group to assess from a statistical perspective the transfer value of professional football (soccer) players. The transfer value is defined as the most likely fee that an engaging team should pay to the releasing club as compensation for breaching the contract of the player wished with respect to the indemnities paid in the past for players with similar characteristics. The presentation describes the sample used to build the econometric model developed, details the variables taken into account, assess both the strength and predictive power of the model. Finally, it introduces some examples of consulting mandates carried out with various football stakeholders. More information is available here: http://www.football-observatory.com/IMG/pdf/note01en.pdf

Keywords

football, transfers, value, price, algorithm

Coaching Athletes with Hidden Disabilities: Collaborative research findings and measurement

Robbi Beyer

Cal State LA, United States rbeyer@calstatela.edu

Abstract. Sport psychology is a field that has recently become more main-stream within Northern American athletics. However, the concepts with the sport psychology field are applicable to many fields within kinesiology, including adapted physical education. Merging these two fields has allowed for improved evaluation, instruction and outcomes for children with hidden disabilities. Hidden disabilities (HD) are defined as disabilities that are not visible or apparent and often aren't detected until the child is given certain learning tasks (Bever, Flores, & Tonsing-Vargas, 2009). Within the athletic arena, young athletes with HD may have trouble following directions, generating strategies, staying focused, and/or getting along with teammates. Recent research has found that in general, youth sport coaches are not prepared to work with athletes with HD, nor are they taught the appropriate pedagogical skills to effectively teach all athletes. Specifically, coaches tend to rely on punitive measures for athletes displaying characteristics of HD and fail to consider their instructional methods as a barrier to learning (Vargas, Beyer & Flores, in 2015). Parents of athletes with HD report different coaching behaviors towards their child (Vargas, Flores, Beyer & Weaver, in press). This can influence the physical progress and psychological well-being of these athletes and although including this topic in coaching education is considered important by coaching educators and administrators, it is scarce. Therefore, the purpose of this presentation is to illustrate the natural fit and collaboration between sport psychology and adapted physical education as it applies to improved coaching education. Focus will be given to current coaching research and measurement that combines the fields of sport psychology, adapted physical education, and special education. This is one of four presentations addressing addressing sport psychology and measurement.

Keywords

sport psychology, coaching, youth sport, adapted physical education

Using a multi-camera tracking system to estimate ball spin in tennis

Olivia Cant^{1,2}, Stephanie Kovalchik^{1,2}, Rod Cross³, and Machar Reid²

¹ Using a multi-camera tracking system to estimate ball spin in tennis, Australia

- ² Game Insight Group, Tennis Australia
- ³ School of Physics, University of Sydney, Australia olivia.cant@live.vu.edu.au christophe.ley@ugent.be groll@statistik.tu-dortmund.de gunther.schauberger@tum.de

Abstract. This study examined the accuracy of methods to estimate ball spin rate and spin axis direction from multi-camera ball tracking technology. Ball spin estimates from (1) the current method implemented by Hawk-Eye (Hawk-Eye Innovations Ltd, Basingstoke, UK) and (2) a theoretical ball trajectory model applied to ball tracking data, were compared for their accuracy to high-speed vision (the ground truth). Ball spin axis direction (topspin, backspin) was estimated with high accuracy by both methods examined, > 98% of trials correctly classified. Ball spin rates were estimated with higher accuracy by the ball trajectory model (RMSE: 222.13 RPM) compared to Hawk-Eye's current ball spin measure (RMSE: 549.56 RPM). A precise and non-invasive measure of ball spin in tennis allows for the collection of match play spin rates, creating the opportunity to investigate the influence of ball spin on player performance using large scale data-sets.

Keywords

ball spin, tennis, tracking technology

Integrated Break and Carryover Minimization Problem in Round Robin Tournaments

Burak Cavdaroglu¹ and Tankut Atan²

¹ Kadir Has University, Turkey

² Bahçeşehir University burak.cavdaroglu@khas.edu.tr sabritankut.atan@eng.bau.edu.tr http://www.khas.edu.tr/cv/2685

Abstract. League scheduling is a field of operations research that has attracted scientists for many years. Break minimization and carryover effect minimization are considered to be two important criteria of fairness in league scheduling. There have been recent studies that integrate both criteria in a computationally hard problem. Some of these studies try to minimize the carryover effect in tournaments so that the number of breaks does not exceed a specific level, while some others apply schedule-then-break approach which first schedule the teams ignoring home-away requirements, then determine the home-away pattern for each team. In this work, we develop a solution method for this integrated problem which produces comparable results with that of a recent study. We show that our method drastically improves carryover effects value at the expense of an occasional increase in the number of breaks.

Keywords

break minimization, carryover effect, league scheduling, round robin tournaments

Modelling the outcomes of professional snooker matches

James Collingwood, Michael Wright, and Roger Brooks

Lancaster University, United Kingdom j.collingwood@lancaster.ac.uk m.wright@lancaster.ac.uk roger.brooks@lancaster.ac.uk https://www.lancaster.ac.uk/lums/people/james-collingwood

Abstract. My research is concerned with how professional snooker matches are won and lost and what differentiates the top players from the rest. My analysis so far has considered different methods of ranking and rating players, ranging from the official World Rankings and player win percentages, to more mathematical models such as Bradley-Terry and Elo. The results from these (which are surprisingly consistent with one another) have enabled me to establish a method of quantifying the relative ability of different players which can be used to estimate the underlying probability that one player will beat another. I have then looked at the extent to which other factors such as current form affects results, and how the chances of winning changes as the match progresses. I have also started gathering performance data from within matches, reviewing the statistics currently produced and considering what other information would be valuable in explaining how matches (and individual frames) are won and lost and why some players are more successful than others.

Keywords

Snooker, Ranking and Rating players, Current Form, Match Progression, Performance Measures

Overcoming the incentive incompatibility of tournaments with multiple group stages

Laszlo Csato

MTA SZTAKI and Corvinus University of Budapest, Hungary laszlo.csato@uni-corvinus.hu https://sites.google.com/site/laszlocsato87/

Abstract. The paper discusses the incentive incompatibility of tournaments with multiple group stages. This design divides the competitors into round-robin groups in the preliminary and main rounds. The higher ranked teams from the preliminary round qualify to the next stage such that matches are not repeated in the main round if two qualified teams have already faced each other. It is proved that these tournament systems, widely used in handball and other sports, violate strategy-proofness since the contestants prefer to carry over better results to the main round. Some historical examples are presented where a team was ex ante disinterested in winning by a high margin. We suggest two incentive compatible mechanisms and compare them with the original format via simulations. Carrying over half of the points scored in the preliminary round turns out to be a promising policy.

Keywords

OR in sports, tournament design, strategy-proofness, simulation, handball

Comparison between some European football leagues through related count data variables

Valentina Cueva, José Rodríguez-Avi, and María José Olmo-Jiménez

Universidad de Jaén, Spain vcueva@ujaen.es jravi@ujaen.es mjolmo@ujaen.es

Abstract. Football is, probably, the most popular sport in Europe, in terms of supporters, news generated, public interest and movement of funds, among others. Weekly, hundreds of thousands of people go to the stadiums, and millions see it by TV, radio, Internet. Even, the number of betting offices increases and all the aspects related to football are especially important.

From a statistical point of view, football can also be seen as a generator of statistical variables that may be studied. In this work, we want to focus on some count data variables that reflect many of the most outstanding football aspects: Specifically, number of goals scored by the player, number of yellow cards and number of cards received by the player. These variables have also been collected in some of the most important football leagues in Europe: Spain, England, Germany and France, and for several years. For each variable, we have proposed several count data model, such as the Negative binomial distribution, the Univariate Generalized Waring Distribution, the Extended Waring distribution or the Complex biparametric and triparametric Pearson distributions, among others. In each case, we select the best fit, in terms of the Akaike Information Criteria or the chi-square goodness of fit test. We compare the results obtained between the different leagues and, for the number of goals scored, we propose regression models that explain this variable in terms of some covariates, such as the number of matches played, the role on the field, the team quality and so on.

Keywords

Football, Extended Waring, Negative binomial, Complex distribution

Randomness of Play Calling in College Football

Brian Curley, Gretchen Hopkirk, Ryan Lokhorst, and Annette Pilkington

University of Notre Dame, United States bcurley@nd.edu ghopkirk@nd.edu rlokhors@nd.edu Anne.B.Pilkington.4@nd.edu

Abstract. In American Football it is widely accepted that the ability to predict the next play of the offensive team gives an advantage to their opponent. The purpose of this study was to examine the effect of a random versus a non-random (and hence more predictable) sequence of offensive plays on outcomes in American College Football. Surprisingly, differences in outcomes were found only for away teams and in this case, the outcomes were seen to depend on the nature of the non-random pattern of play. We saw that non-random play with very few switches between runs of rushing and passing plays did not significantly reduce the chances of winning for either team. In fact, away teams having this style of play showed a significant advantage in some key game statistics. Away teams with non-random play that had many switches between rushing and passing plays had a significantly lower proportion of wins and a significant disadvantage in some game statistics. Our conclusions are that predictability itself does not lead to a serious disadvantage on the playing field in American College Football, rather away teams who exhibit a particular type of non-random play involving many switches between running and passing plays are at a disadvantage. The asymmetry between the results for home and away teams suggest that this type of play plays some role in home advantage. Since play calling is, to a large degree, a factor under the control of the offensive team, knowledge of this association may be useful in forming a strategy for the away team.

Keywords

College football, Playcalling, Game theory

A Modern Love Story: Machine Learning and the Global Sports Betting Industry

Lloyd Danzig

ICED(AI), United States lloyd@icedai.com http://icedai.com

Abstract. Developments in Artificial Intelligence, particularly those within the subfield of Machine Learning, are revolutionizing virtually every industry on the planet. The global sports betting industry, especially with the United States' repeal of PASPA in 2018, is ripe for disruption. Exponential increases in the ability to collect, distribute, and analyze sports data have led to an influx of top engineers entering the space. This presentation will focus on some of the principal ways in which Machine Learning is revolutionizing the industry, ranging from oddsmaking and risk management to fraud detection and responsible gaming implementations. It will also offer a guide to the economics of the business side of the industry and discuss relevant current topics in the tech space, such as adversarial machine learning.

Keywords

sports betting, machine learning, artificial intelligence, risk management

Establishing a performance edge in P2P betting

Carl Donovan¹, Bruno Caneco², and Colin Bleak¹

¹ University of St Andrews, United Kingdom

² DMP Statistical Solutions UK Ltd, United Kingdom crd2@st-andrews.ac.uk bruno@dmpstats.com cb211@st-andrews.ac.uk http://www.dmpstats.com

Abstract. Sports betting is increasingly conducted over peer-to-peer (P2P) platforms such as Betfair or Betdaq. Unlike traditional sports gambling, where participants typically purchase odds from book-makers, online P2P platforms make it equally easy to sell or buy odds. In gambling parlance, participants operate as both bookmakers and punters by being able to take either lay or back positions.

Gamblers are notorious for developing "systems" that give a purported advantage in a betting market which is statistically a positive expectation, post fees or commissions. The systems may be based on theory, data, personal beliefs or mixes thereof, often being tested through actual trading. P2P bet-trading can be highly volatile, and it can be difficult to separate a purported edge from chance occurrence - exacerbated by the ability to take lay positions.

We illustrate here the remarkable ease of being misled by noise in lay positions and the generally low power for detecting genuine edges. We also present here a set of fundamental statistical approaches for assessing the presence and magnitude of gambling edges. In particular, testing for the presence of significant edges, power calculations based on mixture distributions or simulations and related Gambler's Ruin calculations. The exemplars are based on large volumes of fine-scale historical P2P data and trades.

Keywords

sports betting, P2P gambling, sports gambling, gambler's ruin, power

Complex 1 in Male Volleyball as a Markov Chain

Sotiris Drikos

School of Physical Education & Sport Science, University of Athens, Greece sdrikos@gmail.com http://sdrikos.com/

Abstract. In Volleyball, complex 1 consists of the serve's pass (reception) - setting - attack skills in this specified order. This sequence is a stable pattern to win a point. Furthermore, it is important for the teams' success. Taking into account that this pattern is a first-order Markov chain, the creation of a probability transition matrix is feasible. Assuming multinomial likelihood with a Dirichlet prior on the transition probabilities a Markovian transition matrix can be constructed and the calculation of conditional success probabilities is, thus, achievable. Data from the performance analysis of the winning team from recent world championships in three age categories (U19, U21, Men) of male Volleyball is used. Evaluation of the pass through a six-level ordinal scale is possible after the validation of the entire scale. The findings lead to redefining target pass area and to shrinking the evaluation scale at least for the teams under study. Moreover, pass accuracy is necessary because it must give at least two options for attack, but not sufficient condition for the success of attack in all age categories for male Volleyball. In the U19 age category, there is a lack of stabilization in the complex 1 sequence after pass against jump spin serve.

Keywords

Volleyball, reception, age categories, Bayesian analysis, conditional probabilities

Modelling volleyball data using a Bayesian approach

Leonardo Egidi¹ and Ioannis Ntzoufras²

- ¹ Department of Economics, Business, Mathematics and Statistics. University of Trieste, Italy
- ² Department of Statistics. Athens University of Economics and Business, Italy legidi@units.it ntzoufras@aueb.gr http://www.leonardoegidi.com http://www.stat-athens.aueb.gr/~jbn

Abstract. Unlike what happens for other major sports such as football, basketball and baseball, modeling volleyball match outcomes has not been thoroughly addressed by statisticians and mathematicians. The main reason could be the game complexity: the total number of sets is a random variable which ranges from a minimum of three to a maximum of five; the number of points achieved by the two competing teams in each set varies depending on whether they are playing the fifth set or not; the number of final set points for two competing teams may exceed 25 when both the teams reach 24 points (24-deuce). We propose a Bayesian negative binomial model for the points achieved by the team loosing the single set, modelling the probability to realize a point via some team-specific point abilities; the probability of winning a set depends on team specific set abilities. Both point and set abilities are assigned some weakly informative prior distributions. We used goodness of fit tools to compare our proposal with other competing models on Italian Superlega 2017-2018, and MCMC replications from the predictive distribution as a simulation device to reconstruct the league. Preliminary results show that our model outperforms Poisson and binomial models in terms of DIC and is adequate in replicating the final ranking of the league.

Keywords

Volleyball match outcomes, Bayesian negative binomial model, Point and set abilities, Goodness of fit, Prediction

Extreme value prediction: an application to sport records

Giovanni Fonseca¹ and Federica Giummolé²

- ¹ University of Udine, Italy
- ² Ca'Foscari University Venice, Italy giovanni.fonseca@uniud.it giummole@unive.it

Abstract. Extreme value theory studies the extreme deviations from the central portion of a probability distribution. Results in this field have considerable importance in assessing the risk that characterises rare events, such as collapse of the stock market, or earthquakes of exceptional intensity, or floods. In the last years, application of extreme value theory for prediction of sport records have received increased interest by the scientific community. In this work we face the problem of constructing prediction limits for series of extreme values coming from sport data. We propose the use of a calibration procedure applied to the Generalised Extreme Value distribution, in order to obtain a proper predictive distribution for future records. First, a simulation study is performed for comparing the calibrated approach to the estimative one applied to the Generalised Extreme Value distribution and the Generalised Pareto distribution. The results show the improvement of the calibrated prediction procedure on the estimative approach. Secondly, calibrated procedures are applied to series of real data related to sport records. In particular, we consider sequences of annual maxima for different athletic events. Using the proposed calibrated predictive distribution, we show how to predict the exact probability of future records and we discuss the existence and prediction of ultimate records.

Keywords

Athletics, Extreme value, Prediction, Records

Player impact measures for scoring in ice hockey

Carles Sans Fuentes, Niklas Carlsson, and Patrick Lambrix

Linköping University, Sweden csansfuentes@gmail.com niklas.carlsson@liu.se Patrick.Lambrix@liu.se http://www.ida.liu.se/ nikca89/ http://www.ida.liu.se/ patla00/

Abstract. A commonly used method to evaluate player performance is to attribute values to the different actions that players perform and sum up these values every time a player performs these actions. In ice hockey, such metrics include the number of goals, assists, points, plus-minus statistics and recently Corsi and Fenwick. However, these metrics do not capture the context of player actions and the impact they have on the outcome of later actions. Therefore, recent works have introduced more advanced metrics that take into account the context of the actions and perform look-ahead. The use of look-ahead is particularly valuable in low-scoring sports such as ice hockey. In this paper, we first extend a recent approach for measuring a player's impact on a team's scoring. The approach considers the game state and context of each play action, and uses the play sequences that lead to a goal (or other outcome) to learn action-value Q-functions for each state using reinforcement learning. The impact of an action is then calculated as the difference of the Q-values of the state in which the action occurs and its resulting state. In addition to the sum of the impacts due to the player's own actions, we also evaluate a measure in which the player is attributed impacts for all actions performed when the player is on the ice, as well as time-normalized variants. Second, using NHL play-by-play data for several regular seasons, we analyze and compare these and other traditional measures of player impact. All measures satisfy the basic eye-test regarding top players in the league, show high correlation with the salary of both forwards and defenders, and the different quartile-ranges considered are stable. Third, we show that none of the metrics provides a good predictor for the impact that a player will have the next game (e.g., using the previous five games to predict the player's impact in the next game). This helps explain why scoring streaks are heavy-tailed and hard to predict.

Keywords

Performance measures and models, Ice hockey, Sports Analytics

Predicting match outcome in professional Dutch football using tactical performance metrics computed from position tracking data

Floris Goes, Matthias Kempe, and Koen Lemmink

University of Groningen, University Medical Center Groningen, Center for Human Movement Sciences

f.r.goes@umcg.nl
m.kempe@umcg.nl
k.a.p.m.lemmink@umcg.nl

Abstract. Quality as well as quantity of tracking data have rapidly increased over the recent years, and multiple leagues have programs for league-wide collection of tracking data. Tracking data enables in-depth performance analysis, especially with regard to tactics. This already resulted in the development of several Key Performance Indicators (KPI's) related to scoring opportunities, outplaying defenders, numerical balance and territorial advantage. Although some of these KPI's have gained popularity in the analytics community, little research has been conducted to support the link with performance. Therefore, we aim to study the relationship between match outcome and tactical KPI's derived from tracking data. Our dataset contains tracking data of all players and the ball, and match outcome, for 118 Dutch premier league matches. Using tracking data, we identified 72.989 passes. For every pass-reception window we computed KPI's related to numerical superiority, outplayed defenders, territorial gains and scoring opportunities using position data. This individual data was then aggregated over a full match. We then split the dataset in a train and test set, and predicted match outcome using different combinations of features in a logistic regression model. KPI's related to a combination of off-the-ball features seemed to be the best predictor of match outcome (accuracy of 64.0% and a log loss of 0.67), followed by KPI's related to the creation of scoring opportunities (accuracy of 58% and a log loss of 0.69). This indicates that although most (commercially) available KPI's are based on ball-events, the most important information seems to be in off-the-ball activity. We have demonstrated that tactical KPI's computed from tracking data are relatively good predictors of match outcome. As off-the-ball activity seems to be the main predictor of match outcome, tracking data seems to provide much more insight than notational analysis.

Keywords

Soccer, Spatiotemporal analysis, Tactical behaviour, Performance analysis, Tracking data

Champions League or domestic league: a coach's choice

Dries Goosens¹, Chang Wang², and Martina Vandebroek²

² Katholieke Universiteit Leuven, Belgium dries.goossens@ugent.be chang.wang@kuleuven.be Martina.Vandebroek@kuleuven.be

Abstract. With the increase of the number of football matches, at both club and national team level, match schedules became more and more congested, and professional football players are increasingly exposed to fatigue and injuries. In order to keep players at a satisfactory fitness level, club coaches have a clear incentive to make player rotations in their line-up during the season. They need to decide in which matches to line-up their key players, and consequently on which competitions to focus.

In this contribution, we develop a new measure to quantify the relative commitment teams show for domestic and European competitions (i.e., UEFA Champions League and UEFA Europa League). By studying player rotation and the quality of their line-up in the domestic league match before and after the European match, we investigate on which competition clubs from ten main associations in Europe (Spain, Germany, England, Italy, France, Portugal, Russia, Ukraine, The Netherlands and Belgium) focus. As far as we are aware, no such measure has been described in the literature before.

We discuss the main differences with respect to commitment to the European competitions between these countries, and we study the matches that deviate from the common strategy. Finally, we also study the impact on performance in the domestic league of the choices with respect to the player line-up.

Keywords

football commitment, player rotation, line-up

¹ Ghent University, Belgium

Play-by-play data analysis for team managing in basketball

Luca Grassetti, Ruggero Bellio, Giovanni Fonseca, and Paolo Vidoni

University of Udine, Italy luca.grassetti@uniud.it ruggero.bellio@uniud.it giovanni.fonseca@uniud.it paolo.vidoni@uniud.it

Abstract. The sport analytics literature regarding basketball is vast but the analyses based on disaggregated data, such as the play-by-play match data, are not well developed. The analysis of the whole sequence of play-by-play match events has an undeveloped potential, yet most of the available methods focus on the final match results. The present work aims at finding a modelbased strategy for the analysis of the match progress, which may support the decision-making process of the team staff and, to a lesser extent, be used for the prediction of match results. The main idea is to describe the outcome of the observed plays as a function of the efficiency of the players on the field and of some contextual variables. The latter may include, for instance, the period of the match or home effect. For the empirical analysis, we consider a dataset regarding the Italian Basketball League (Serie A1). At first, the matches of the first round of the current championship 2018/2019 are considered. The dataset collects the play-by-play information along with the matches box scores, which are made available by the league website (www.legabasket.it). The plays are then aggregated by "shifts", defined as periods between subsequent substitutions, in order to consider a more manageable outcome measure. Suitable regularized regression models are then adopted to estimate main player effects along with 5-man unit effects. The methodology can be used for various aims, including the evaluation of player efficiency which can be compared to existing adjusted plus-minus measures. A further application is to suggest some guidance on whether alternative match strategies could have been adopted with better performances.

Keywords

Basketball Analytics, Statistical Model, Play-by-play data, Web-crawling, Data-driven decision process,

Fair-Fixture: Minimizing Carry-Over Effects In Football Leagues

Dilek Günneç¹ and Ezgi Demir²

- ¹ Department of Industrial Engineering, Özyeğin University, Istanbul, 34794, Turkey dilek.gunnec@ozyegin.edu.tr
- ² BSH Home Appliances Corporation, Istanbul, 34771, Turkey

Abstract. We study a sports scheduling problem with the objective of minimizing carry-over effects in round robin tournaments. In the first part, focusing on tournaments that allow minimum number of breaks (at most one) for each team, we formulate an integer programming model and provide an efficient heuristic algorithm to solve this computationally expensive problem. We apply the algorithm to the current Turkish Professional Football League and present an alternative scheduling template. In the second part, we discuss how the carry-over effects can be further decreased if the number of breaks is allowed to be of slightly larger value and numerically represent this trade-off.

Will Groups of 3 Ruin the World Cup?

Julien Guyon

Columbia University and New York University, United States julien.guyon.1977@gmail.com http://cermics.enpc.fr/ guyon/

Abstract. In 2026, and maybe even as soon as 2022, the FIFA World Cup will for the first time gather 48 men's national teams. It will consist of a group stage made of 16 groups of three, with the best two teams in each group advancing to the knockout stage. Using groups of three raises several fairness issues, including risk of match fixing and schedule imbalance. In this article we examine the risk of collusion. The two teams who play the last game in the group know exactly what results will let them advance to the knockout stage. Suspicion of match fixing occurs when a result qualifies both of them at the expense of the third team of the group, and can seriously tarnish the tournament. We quantify how often this is expected to happen and explain how to build the match schedule so as to minimize the risk of collusion. We also quantify how the risk of collusion depends on competitive balance. Moreover, we show that forbidding draws during the group stage (a rule considered by FIFA) does not eliminate the risk of match fixing, and that surprisingly when draws are forbidden the 3-2-1-0 point system does not do a better job at decreasing the risk of collusion than the 3-0 point system. Finally we describe alternate formats for a 48 team World Cup that would eliminate or strongly decrease the risk of collusion. This work has been published in The New York Times [1] and in Le Monde [2]. The New York Times also published a follow-up article [3].

[1] Guyon, J.: Why Groups of 3 Will Ruin the World Cup (So Enjoy This One), The New York Times, June 11, 2018.

[2] Guyon, J.: Mondial 2026 : pourquoi les groupes de trois risquent de fausser la Coupe du monde, Le Monde, June 12, 2018.

[3] Guyon, J. and Monkovic, T.: FIFA, We Fixed Your World Cup Collusion Problem for You, The New York Times, June 26, 2018.

Keywords

tournament design, scheduling, fairness, FIFA World Cup, groups of 3, match fixing, collusion

Analysing the effect of a change of transition probabilities related to possession on scoring a goal in a football match

Nobuyoshi Hirotsu¹ and Ayako Komine²

 Juntendo University, Japan
 Japan Sport Council n.hirotsu@sakura.juntendo.ac.jp

Abstract. In this paper, we use a twenty-state Markov process model of a football match to analyse the effect of a change of transition probabilities related to possession on scoring a goal. In the model, we divide the pitch into 9 areas, and collect the data in terms of the change of location of the ball, together with the change of possession of the ball. A log-linear model, fed by annual data from the J League Division 1 in 2015, is used to estimate the transition probabilities by means of the maximum likelihood method. We make a change of the strength of a team in terms of possession which affects the transition probabilities. Using these transition probabilities, we obtain the probability distribution of scoring a goal, and analyse the effect of the change of transition probabilities on scoring a goal. Practically, this change can be realized by tactical changes such as substitution of players. A simple example of this approach is included as an illustration.

Keywords

Football match, J League, Markov process, Possession of the ball, Scoring a goal

Score-based soccer match outcome modelling - an experimental review

Ondřej Hubáček, Gustav Šourek, and Filip Železný

Czech Technical University in Prague hubacon2@fel.cvut.cz souregus@fel.cvut.cz zelezny@fel.cvut.cz

Abstract. We investigate the state-of-the-art in score-based soccer match outcome prediction modeling to identify the top-performing methods across the diverse classes of existing approaches to the problem. Namely, we bring together methods ranging from (i) statistical models, based on popular distributions such as Poisson and Weibull, (ii) general ranking algorithms, such as the famous Elo, together with the more contemporary representatives, e.g. Trueskill, (iii) domain-specific rating systems such as pi-ratings, and advanced machine learning methods utilizing (iv) structured, i.e. graph-based, representation of the problem (v) as well as a proven ensemble model based on carefully designed expert features.

We investigate assumptions, complexity, advantages and behavior of the studied methods to provide further insights and draw connections across the, typically separately studied, approaches to the problem. Finally, we experimentally compare these diverse competitors altogether on a large database of soccer results to identify the true leaders across different evaluation measures.

Keywords

forecasting, soccer, experimental review

Modelling the 2018 FIFA World Cup - Could we have "Beaten the Bookies" ?

Gordon Hunter¹, Benjamin Jauvion², Ishan Rashid¹, and Mohamed Sharif-Ali¹

¹ Kingston University, United Kingdom

² Polytech Clermont-Ferrand, Université Clermont Auvergne G.Hunter@kingston.ac.uk benjamin.jauvion@etu.uca.fr ishan_rashid@btconnect.com sharifali_92@hotmail.co.uk http://www.kingston.ac.uk/staff/profile/dr-gordon-hunter-73/

Abstract. Gambling on the results of football matches has become very popular over recent years, and indeed one sponsor offered a GBP ÂL100 million prize to anyone who could correctly predict the outcomes of all 64 games in the 2018 FIFA World Cup. No-one won that prize, but could a "smart punter" using a good model have made a substantial profit by placing conventional bets on the results of those World Cup matches ?

In this paper, we describe our attempts to model the matches of the 2018 FIFA World Cup using a variety of approaches. Some of our models were adapted from Maher's (1982) classic model of football scores, others based on multinomial logistic regression using known attributes (e.g. recent record, FIFA ranking) of the teams playing each game. We then used the results of our models, in conjunction a variety of gambling strategies (including variants of Kelly betting) to place virtual bets on the outcome of each match, and investigated our (virtual) return on investment in each case based on the actual match results and pre-match bookmakers' odds. Some models and strategies did better than others, but some did provide a positive net return for the actual set of 2018 World Cup match results. (Our work was being carried out whilst the 2018 World Cup was actually taking place.)

Whilst it is over-ambitious to claim that our models and strategies could enable us to "beat the bookies" in a general situation, we were able to do so for this particular set of results.

Keywords

Football, Soccer, Gambling, Regression Model, Maher Model, Poisson Distribution, Maximum Likelihood Estimation, Simulation, Kelly Betting
Using assessments to facilitate mental performance consulting

Madison Hunt and Anastasia Blevins

Long Beach State University, United States leilani.madrigal@csulb.edu Anastasia.Blevins@student.csulb.edu

Abstract. In a recent study by the Sports and Fitness Industry Association and the Aspen Institute, youth sport participation in the USA has declined by almost 8 percent for the "big four"t youth sports: baseball, basketball, football, and soccer between 2008-2016 (Bogage, 2017). For many young people, sport participation that starts in childhood stops in adolescence, often due to perceived competence, autonomy, relatedness, and task climate (Balish, McLaren, Rainham, & Blanchard, 2014). Additional factors such as lack of enjoyment, anxiety due to excessive criticism, injury, and perceptions of negative team dynamics have also contributed to reasons youth drop out of sports (Witt & Dangi, 2018). But this decision may also be due, in part, to athletes' mental skills use. Assessments in sport psychology are useful to help guide the research to practice. Such assessments can also help guide sport health care professionals in how they interact and work with athletes. In a recent study with adolescent club volleyball athletes, the Athletic Skills Coping Inventory (ACSI) was used to assess the athletes' psychological skills pre-season. The ACSI measures coping with adversity, coachability, concentration, confidence and achievement motivation, goal setting and mental preparation, peaking under pressure and freedom from worry. This information then helps guide the consultant over the course of the season in what mental skills to teach. This information, in combination informal observation of practices, dictated the skills taught to athletes. Over the course of the competitive season, athletes reported increases in several of the ACSI subscales. Additional strategies on how consultants can use assessments with youth for consulting and other populations will be presented. This presentation will serve as part of the talks on sport psychology.

Keywords

mental skills, consultant, psychological assessment

Serve and Return Glicko: An approximate Bayesian model for within-match win prediction

Martin Ingram

The University of Melbourne, Australia ingramm@student.unimelb.edu.au

Abstract. Within-match forecasting is a challenging task in tennis. On the one hand, information about each player's form on the day is revealed over the course of the match; on the other hand, this information must be balanced against the possibility of changes in skill during the match caused, for example, by fatigue.

Building on the dynamic paired comparison model Glicko, I investigate the use of fitting separate serve and return ratings over time to model player skill over time. Within a match, these ratings are updated dynamically according to point outcomes; between matches, the variance around each player's skill estimates is assumed to grow in proportion to the time since their last match, reflecting growing uncertainty. The resulting model can be used both for within-match prediction as well as pre-match forecasting. Using data from the ATP and WTA tours, I evaluate the model against other approaches to within-match updating as well as against other pre-match prediction models and interpret the inferences from the model. I analyse to what extent the model suggests that momentum swings exist in a match and discuss what implications this has for within-match forecasting in the sport.

Keywords

rating systems, paired comparison, Bayesian inference

Sports Coaching: An Art and a Science

Simon Jenkins

University of Winchester, United Kingdom simon.jenkins@winchester.ac.uk

Abstract. Sports coaching is both an art and a science, but what is meant by 'art' and 'science'? What expertise and experience does a coach require? How can a university education be of benefit to a sports coach? What makes a great coach and why is John Wooden so revered? These questions are addressed by providing: i) critical insight on how sports degrees in the UK have changed since the 1980s; ii) an overview of research and scholarship in sports coaching; and iii) reference to recent discussions about 'professionalisation' (i.e., occupational status and standing) and 'professionalism' (i.e., quality and standards).

Keywords

Aristotle, Coaching philosophy, Coach John Wooden, Critical thinking, Evidence-based practice, Expertise, Golf, Paradigm, Performance Analysis, Philosophy of science, Phronesis, Professionalism, Reflective practice, Scholarship, Scientism, Specialised knowledge, Sports degrees, Statistics, Weightlifting

The explosive growth of eSports and the potential for research opportunities

Lyn Kee, Minh Huynh, Denny Meyer, and Kelly Marshall

Swinburne University of Technology, Australia ykee@swin.edu.au mhuynh@swin.edu.au dmeyer@swin.edu.au kmmarshall@swin.edu.au

Abstract. It is estimated that over 300 million people worldwide watch eSports events, both live and online. This viewership shows no sign of slowing down and is expected to grow by 12% each year. The growing popularity of eSports has thus transformed the industry into an excellent revenue-making opportunity. According to market research, the global revenue of eSports was over \$1.5 billion in 2017 alone. Consequentially, the opportunity for such a growing industry has attracted a wide range of investment opportunities. For example, applications of eSports within traditional sporting contexts, such as modelling and betting, are leading such applications and are one of the most prolific areas. According to Pinnacle Sports, one of the eSports betting leaders, betting on eSports events have surged from 100 thousand to five million only a five-year period. Given that research in eSports betting and modelling is still in its infancy, it warrants a promising research potential. This paper will outline the relatively untapped eSports betting industry and discuss opportunities for researchers to apply statistical methods and to collaborate within this growing field.

Keywords

eSports, modelling, betting

Move it or lose it: Exploring the relation of defensive disruptiveness and team success

Matthias Kempe and Floris Goes

University of Groningen, University Medical Center Groningen, Center for Human Movement Sciences, The Netherlands m.kempe@umcg.nl f.r.goes@umcg.nl

Abstract. Introduction: Due to the increasing number of tracking data available for official matches in different leagues there are new ways to capture the performance of teams. To not rely on notational data, we previously introduced the D-Def (Goes et. al, 2018), an aggregated variable to quantify passing solely based on tracking data. This value captures the change of organisation by a pass (defensive disruptiveness). In this study, we updated the D-Def by including an automated classifier for subunits, instead of using starting formations, and investigated the relation of the D-Def on team success.

Methods: Position tracking data of all players and the ball collected during 89 Dutch Premier League matches was used. Alignment of subunits was automatically identified, using a K-Means classifier, for every pass. D-Def was calculated for every pass (N = 63601) as an aggregate in the change in movement as a result of the pass-based team- and line centroids of subunits and surface and spread of the defending team. Team success was evaluated via wins and losses. We excluded 21 matches because they resulted in a draw. The predictive value of the D-Def for success was calculated using logistic regression analysis.

Results & Discussion: The regression model achieved a RÅ of 0.69 which is high in comparison to other key performance indicators in the literature. This shows that the approach previously introduced as a proof of concept is related to match outcome. Therefore D-Def can be a useful tool to evaluate team performance. Conclusions: This study highlights that performance can be predicted through spatio-temporal aggregates based on player tracking data and we do not need to rely on notational data anymore.

Reference:

Goes FR, Kempe M, Meerhoff LA, Lemmink KAPM (2018). Not every pass can be an assist: a data-driven model to measure pass effectiveness in professional soccer matches. Big Data, DOI: 10.1089/big.2018.0067

Keywords

Analytics, Performance Analysis, Soccer, Big Data

An evaluation of the three-point rule in football

Wim Koevoets

Individual, Portugal ger4ec@gmail.com

Abstract. The three-point rule in association football replaced the two-point rule by awarding three points to teams that win a match instead of two. The Isthmian league introduced this rule in 1973 to make football more attractive to watch. Most national football associations implemented the rule in their competitions after the 1994 World Cup.

The widespread adoption of the three-point rule suggests that the associations are better off with this rule than they would have been with the two-point rule. We investigate this suggestion by describing general mechanisms through which we expect an effect of the introduction of the three-point rule on matches to operate.

We use information on goals to estimate the relation between the introduction of the threepoint rule and match outcomes for matches in the Italian Serie A. The qualitative model and Maher (1982) form the basis of our statistical models. If you belief the qualitative model then you can interpret the estimated relation as a causal effect of the introduction of the three-point rule.

Keywords

Football, Three-point rule, Incentives, Directed acyclic graph, Evaluation, Causal effect, Poisson, Logit, Weibull

A quantitative method for evaluating the skills of national volleyball teams: Prediction accuracy comparisons of the official ranking system in the worldwide tournaments of 2010s

Eiji Konaka

Meijo University, Japan konaka@meijo-u.ac.jp

Abstract. This paper proposes a quantitative skill-evaluation method for international volleyball teams. The main objective of this paper is to point out design flaws in the official FIVB (Fédération Internationale de Volleyball Association) ranking by comparing its predictions of performances with the proposed method in major worldwide tournaments (e.g., World Championships and Olympic Games, held in the 2010s). In the proposed method, only one parameter, rating, shows the scoring skill of each team. Assume that the difference in rating values explains the scoring ratio in one match via a logistic regression model. The rating values are calculated by an iterative method to minimize squared error between the predicted and actual scoring ratio. Match result prediction was performed for ten tournaments and 733 matches. For each tournament, its qualification tournaments, worldwide leagues, and continental championships within about two years was used to evaluate skills. The rating values were calculated using over 8,000 international match results. As a result, prediction accuracy based on the FIVB ranking and the proposed method were 0.742 and 0.771, respectively. The proposed method can improve prediction accuracy, but the difference does not show statistical significance. Predictions were also made about whether each team could advance to the second round. In the first round, each team was divided into pools, and played in a round-robin system with several top teams in each pool advancing to the second round. The prediction accuracy for all 192 teams based on the FIVB ranking and the proposed method were 0.776 and 0.875, respectively. This difference shows statistically significance. Using a detailed analysis, in the FIVB rankings, European teams were always under-evaluated, while the remaining teams were always over-evaluated. This is clear evidence of the design flaws in the FIVB ranking system. Finally, as a case study, the World Championships in 2018 were analyzed from the viewpoint of all-Japan teams.

Keywords

volleyball, rating, match prediction, ranking system

Availability-to-train in elite sport

Ioannis Kosmidis¹

- ¹ The University of Warwick, United Kingdom
- ² Tesseract Academy, United Kingdom ioannis.kosmidis@warwick.ac.uk http://www.ikosmidis.com

Abstract. Recent technological advances have allowed the frequent and simultaneous measurement at small cost of multiple physiological and psychological aspects of individuals and of their fitness and everyday activity. Such data are increasingly being used, for example by elite sport teams to quantify and monitor athlete fatigue and identify the variables that are important for explaining it, with the aim of preventing non-functional overreaching.

This talk will discuss the statistical challenges involved in recording and preparing fatigueand fitness-related data for statistical modelling, including missingness and measurement errors. We will also present results from ongoing research on modelling availability-to-train in terms of other fatigue and fitness proxies, accounting for not only the heterogeneity between athletes, but also for the natural time dependence between measurements.

Keywords

competition data, performance data, exploratory factor analysis, stochastic gradient descent, validation

Random Walks with Memory Applied to Grand Slam Tennis Matches Modeling

Tomáš Kouřim

Institute of Information Theory and Automation, Czech Academy of Sciences kourim@outlook.com

Abstract. The contribution presents a model of a random walk with varying transition probabilities (Kouřim, 2017) implicitly depending on the entire history of the walk, which is an improvement of a model with varying step sizes (Turban, 2010). The transition probabilities are altered according to the last step of the walker using a parameter to either reward or punish success by increasing or decreasing its probability in the next step.

Such a walk seems to suitably approximate the development of a sport game. This theory is inspected by applying the walk as a model of a tennis match. The focus is on the best-of-five tennis games, i.e. the man Grand Slam tournaments, with each set representing one step of the random walk. The model is fitted on the entire history of matches since 2009 and thoroughly examined.

Further, the theory is extended, and the cases with two parameters (one affecting each player's odds of winning) and with time dependent parameter are studied. Theoretical assumptions are proposed, proved and applied on real data.

Differences between respective tournaments are examined, as are the differences between the individual players. The work especially focuses on the top players from the last decade. Additionally, the matches are grouped using bookmaker's odds into more homogenous subsets and the model is applied to each one separately. Again, results are compared among the groups and within each group as well.

Main sources: Tomáš Kouřim. Random walks with varying transition probabilities. Doktorandské dny FJFI, 2017.

http://kmwww.fjfi.cvut.cz/ddny/historie/17-sbornik.pdf.

Loïc Turban. On a random walk with memory and its relation with markovian processes. Journal of Physics A: Mathematical and Theoretical, 43(28):285006, 2010. https://arxiv.org/pdf/1005.3896.

Keywords

random walk with memory, random processes, tennis match modeling, in-play odds prediction, Grand Slam

Extensions of the Elo Rating System for Margin of Victory

Stephanie Kovalchik¹

Institute of Sport, Exercise and Active Living, Victoria University, Melbourne, Australia Stephanie.Kovalchik@vu.edu.au

Abstract. The Elo rating system is one of the most popular methods for estimating competitor ability over time in sport. The standard Elo system focuses on the prediction of wins and losses but often there is also interest in the margin of victory (MOV) as it reflects the magnitude of a result. Theory for and comparisons of these types of models are limited. The present work proposes four model options for an MOV Elo system: linear, joint additive, multiplicative, and logistic. Notation and guidance for tuning each model is described. The models are applied to men's tennis for several choices of MOV. The results show that all MOV approaches using within-set statistics improve predictive performance over the standard Elo system, while only the joint additive model yielded unbiased ratings with stable variance in a simulation study. This general framework for MOV Elo ratings gives sports modellers a new set of tools when building systems to rate competitors and forecast outcomes in sport.

Fair sudden death

Roel Lambers and Frits Spieksma

Eindhoven University of Technology, Netherlands r.lambers@tue.nl f.c.r.spieksma@tue.nl

Abstract. We consider fairness of penalty shoot-outs given a First-Mover Advantage, with a focus on a sudden death format where the order of shooting is predetermined. The current ABAB-format is known to be unfair, since the team shooting first has a higher probability of winning. The goal of this contribution is to show under which conditions fair penalty shoot-outs exists.

Keywords

Fairness, Shoot-outs, First-mover advantage

On the flexibility of Home-Away pattern sets

Roel Lambers¹, Dries Goossens², and Frits Spieksma¹

¹ Eindhoven University of Technology, Netherlands

² Ghent University, Belgium r.lambers@tue.nl dries.goossens@ugent.be f.c.r.spieksma@tue.nl

Abstract. We consider the scheduling of a sports league in a round robin format. We investigate to what extent the choice of Home-Away patternsets (HAP-sets) leaves room for the assignment of individual matches to rounds. Therefore, we introduce three measures representing the flexibility of HAP-set, and we give explicit expressions for the values of these measures for some well-known choices of the HAP-sets.

Keywords

round robin tournaments, scheduling, Home Away patterns

The Age Advantage in Youth Football

Steve Lawrence¹, Laura Jonker², and Jan Verbeek³

- ¹ Milk Studios, United Kingdom
- ² University of Groningen, Netherlands
- ³ Koninklijke Nederlandse Voetbalbond, Netherlands steve.lawrence.ata@gmail.com laura@xoet.nl jan.verbeek@knvb.nl http://www.miteamsheet.com

Abstract. Average Team Age (ATA) and a Relative Age Index (RAEi) are variables against which performance outcomes in football can be measured and we consistently find that performance advantages are evident when measured against these variables. In 'The Age Advantage in Association Football', Lawrence, S., MSp2015, 6,389 matches played by males at U17 to adult team ages were examined providing evidence of an age advantage. In this new paper analysis of additional match data from U12 to adult, providing a total dataset of 15,088 matches, is presented providing further insight into the development of relative age effects and signifying a causal connection between cut-off date eligibility rules and such effects. We conclude with a proposition that replacement of cut-off date rules with 'average team age; rules will assist with the elimination of such effects.

Keywords

Relative age effect, Relative age bias, Average age, Average team age, Relative age advantage, Age advantage, RAEi, ATA

Survival Modelling of Goal Arrival Times in Champions League

Ilias Leriou, Ioannis Ntzoufras, and Dimitris Karlis

Athens University of Economics and Business, Greece eleriou@aueb.gr ntzoufras@aueb.gr karlis@aueb.gr

Abstract. In this work, we consider possible survival modelling of goal arrival times using the bivariate Weibull distribution under a competing risks framework. The proposed approach takes into account the competitiveness of each team's goal arrival time by considering that the arrival goal time for the scoring team is taken as the censoring time for the opponent. Estimation of the parameters is possible using MCMC methodology. Finally, a final model is selected using Gibbs Variable Selection (Dellaportas, P. et al. 2000) and that final model is compared in terms of its goodness of fit and prediction with existing models, like the Negative Binomial Model and the Poisson Model (Karlis, D & Ntzoufras, I. 2003). The proposed methodology will be presented on data concerning the Champions League 2017-2018.

References: Karlis, D., and Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. Journal of the Royal Statistical Society: Series D (The Statistician), 52(3), 381-393.. Dellaportas, P., Forster, J. J. and Ntzoufras, I. (2000). Bayesian variable selection using the Gibbs sampler. BIOSTATISTICS-BASEL- 5, 273-286

Keywords

Survival, Bivariate Weibull, Goal Arrival Times

Talent detection in sport : Machine Learning methods for performance prediction

Arthur Leroy

Université Paris Descartes — INSEP, France arthur.leroy@parisdescartes.fr http://arthur-leroy.netlify.com

Abstract. Many data collected in sport science come from time dependent phenomena. Several sport structures such as federations or clubs have collected these longitudinal data, hoping it may help the identification of future talents. Moreover, recent studies highlighted that most of top-level young athletes do not remain at the best level of performance at adult age. Thus, the talent identification issue might benefit from data analysis and particularly from machine learning methods. In the present paper, the main goal is to predict the future performance of an athlete, using its own past performances and information from other athletes. Due to their genuine nature and the uneven observation timestamps, performance data are considered as functional over time and fitted thanks to B-spline basis. These functional observations are assumed to be realizations of a Gaussian stochastic process, and the problem is addressed through Gaussian process regression. This task is treated by a mixed effect model where a central process is common to all observations and a random process is fitted on each individual curve. This approach allows to rely on information from the whole dataset for modelling and then settles the issue of sparse observations. A first clustering step is performed on the functional data, which allows a cluster-specific prediction step afterwards. The whole procedure is estimated from a Bayesian perspective, which naturally enables the computing of uncertainty and credible intervals for our predictions. The effectiveness of the method is assessed through a simulation study as well as an application on a real data-set from the French Swimming Federation. The usefulness of the method is twofold, offering a better understanding of the performance progression phenomenon in sport, and providing an automatic talent identification tool.

Keywords

Talent detection, Prediction, Functional Data, Gaussian Process, B-splines

Favorite longshot bias and betting margin: An insider's perspective

Kostas Maras

Stoixman.gr, Greece k.maras@stoiximan.gr

Abstract. One of the most commonly accepted phenomena in the world of sports betting is the favorite-longshot bias. There is a large number of studies proving the longstanding empirical regularity that betting odds provide biased estimates of the winner probability, as the bettors tend to overvalue "long shots" and undervalue favorites, while bookmakers respond to this punter bias by applying disproportionately the respective over-round. Behavioral psychology indicates that punters estimations and decisions are subject to numerous biases and the expansion of Inplay betting intensified such dependencies. On the contrary, we will check out the betting patterns regarding the multiples dominated Pregame subset in Greece and we will identify a reverse favoritelongshot bias hypothesis. We conclude the presentation with a description of the 0% margin paradox and the unexpected illustration of both favorite and reverse favorite long-shot bias on the same event.

Assessing Mental Toughness

Leilani Madrigal

Long Beach State University, United States leilani.madrigal@csulb.edu

Abstract. Shooting a penalty kick during the World Cup. Finishing the final stage in the Tour de France. Persisting through fatigue in an Olympic marathon. Mental toughness has been a primary construct encompassing the psychological attributes many consider necessary for success in sports (Cowden, 2017; Gucciardi, Gordon, & Dimmock, 2009a). It is defined as the "collection of values, attitudes, behaviors, and emotions that enable you to persevere and overcome any obstacle, adversity, or pressure experienced, but also maintain concentration and motivation when things are going well to consistently achieve your goals"t (Gucciardi Gordon, & Dimmock, 2008, p. 278). The concept of mental toughness is particularly relevant in sport psychology research, but related research has been criticized, in part due to measurement inconsistencies (Crust, 2008). Key shortcomings of such measures include general vs sport-specific measures, conceptual models, and multidimensional vs unidimensional measures. This presentation will review some of these shortcomings and introduce a measure of mental toughness that was developed to counter some of these limitations. The Mental Toughness Scale (MTS; Madrigal, Gill, & Hamill, 2013) was developed to address some of the criticisms leveled against previous mental toughness measures; specifically, using the framework proposed by Jones, Hanton, and Connaughton (2007), the authors developed the tool to be generalizable across sports and athlete gender. The MTS has demonstrated excellent psychometric testing through confirmatory factor analyses, gender equivalence, internal consistency, and convergent and divergent validity (Madrigal et al., 2013; Madrigal, Gill, & Willse, 2017; Madrigal, Wurse, & Gill, 2016). The presentation will conclude with additional factors to consider when assessing mental toughness in sport. This presentation will serve as part of the talks on sport psychology.

Keywords

Mental Toughness, Athletics, Measurement

Maximizing financial and on-field performances when composing teams in soccer

Lars Magnus Hvattum¹ and Giovanni Pantuso²

¹ Molde University College, Norway

² University of Copenhagen, Denmark http://www.himolde.no/english/Sider/Minside/Lars.M.Hvattum.aspx

Abstract. Composing a team of players is among the most crucial decisions a soccer club is required to make. Transfer fees and player wages make up a significant portion of the budget for many top clubs, and the sporting success of a club is tightly associated with its income. Despite this, there is relatively little academic research addressing decision support for clubs trying to determine which players to purchase and which players to sell during a transfer window.

This work presents a chance-constrained mixed-integer programming problem that aims to maximize the quality of the players on a team, given the opportunity to buy players, sell players, and enter loan agreements with players during a single transfer window. The team must adhere to certain constraints regarding team composition, based on different player roles and team formations as well as on age restrictions and roster size. In addition, there is a budget constraint, to avoid overspending, and a chance-constraint indicating that the value of the entire team must be above a threshold with a given probability.

Historical data on player transfers are used to build a linear regression model predicting future values for players. A novel plus-minus rating is developed to estimate the quality of individual players. A regularized plus-minus rating is based on using ridge regression to estimate a linear regression model where some of the regression coefficients correspond to the rating of individual players. Our plus-minus rating is a generalization of this, formulating an unconstrained quadratic programming problem that also allows the estimation of ageing effects, as well as for differences in playing strength between different competitions. A computational study is conducted to verify that a club aiming to maximize its sporting performance will make other team composition decisions than a club focusing mainly on maximizing the total value of its team.

Keywords

team composition, integer programming, stochastic programming, prescriptive analytics

Passing tactical relevance during soccer matches determined by machine learning

Murilo Merlin¹, Felipe Moura¹, Ricardo Torres¹, Vitor Principe², and Sergio Augusto Cunha¹

- ¹ State University of Campinas, Brazil
- ² Rio de Janeiro State University, Brazil murilomerlin7@gmail.com felipemoura@uel.br rtorres@ic.unicamp.br vitorprin@gmail.com scunha@fef.unicamp.br

Abstract. Technological and computational advances have challenged coaches and scientists to raise new questions about soccer match, mainly from a multivariate perspective. This scenario has led to an increase of the use of machine learning techniques in different applications. The pass in soccer is the most used technical-tactical action during the match and has been considered an important performance indicator. Usually, the pass is analyzed based on effectiveness (successful or not successful) limiting the quality of information. Classifying the passes based on most significant concept to performance would allow more relevant analysis about the match. Therefore, the aim of this study was to propose an automatic pass classification model in soccer matches based on tactical relevance (TR) concept, from spatiotemporal data. Four official matches of the Brazilian Football Championship 2016 were tracked by video-based system. From those, 33 passing predictor variables grouped in pressure, pitch position, ball trajectory and passing result were proposed. Three soccer experts labelled 465 passes into five classes, very bad (2.6%), bad (12.5%), medium (66.2%), good (17.4%), very good (1.3%), based on TR concept (advantage degree provided by the passing in a given match context). We used random forest (RF) algorithm for passes classification (acikit-learn v0.20.2 library, Python 3.7, stratified k-fold cross-validation). The RF classification model presented accuracy of 80.0% for passes classification to the five classes, similar Horton & Gudmundsson (2014) who obtained accuracy of 85.0% to classify the quality of the pass into three different classes. We proposed the original TR concept, predictor variables based on the opinion of experts, and a classification model based on a nonlinear algorithm. The classification model will allow several practical applications, from player ranking, offensive sequence analysis, variable power testing and training planning.

Keywords

soccer, passing, tactical, machine learning, random forest

Fame and Fortune in Elite Tennis Revisited

Denny Meyer¹, Minh Huynh¹, Kelly Marshall¹, and Geoff Pollard²

¹ Swinburne University of Technology, Australia

² Tennis Australia dmeyer@swin.edu.au mhuynh@swin.edu.au kmmarshall@swin.edu.au GPollard@Tennis.com.au

Abstract. Professional tennis is aging with older players able to prolong their careers thanks to science, technology and healthy pay cheques. However, change is in the air making this a good time to investigate the trends and patterns apparent in ATP statistics. Previous research, using data for only the top 128 ATP players in the 2011 ATP World Tour Media Guide, investigated the relationship between performance, prize money and rankings for the period 2004 to 2010. The results were contradictory, depending on what measure of performance were used. This paper again investigates the relationship between rankings, prize money and performance using On Court data for the top 200 ranked players for the years 2004 to 2018. This larger, more modern data finds more consistent results, with prize money consistently more important than rankings for predicting performance. Secondary analyses consider changes that generally occur through a player's career in terms of prize money, average rankings, percentage of matches and tie breaks won, average number of aces per match and several other important match parameters. Linear trends are detected, suggesting consistent changes over the careers of elite players. However, in the case of matches played, average rankings and prize money, the magnitude of the time trends are dependent on age in 2018, suggesting different trajectories for younger and older generation players. More rapid improvements are seen amongst younger generation than older generation players. The implications for younger players are exciting, suggesting that there will soon be a changing of the guard.

Keywords

rankings, prize money, multi-level models, performance incentives, career trajectories

Flexible multivariate point processes with applications to modelling football matches

Santhosh Narayanan¹, Ioannis Kosmidis¹, and Petros Dellaportas²

¹ The University of Warwick, United Kingdom ² University College London, United Kingdom nsanthosh27@gmail.com ioannis.kosmidis@warwick.ac.uk p.dellaportas@ucl.ac.uk

Abstract. The analysis of football data is mathematically challenging since it is an instance of a spatio-temporal analysis problem with complex, interdependent events and trajectories. Over the last decade, there has been a concerted effort to record the events that happen during the course of a game at high frequency and accuracy. The resulting data is directly relevant in player performance evaluation for coaching and scouting, the development of the game strategies and in enhancing the viewing experience of televised games. We are developing multivariate point processes suitable for the modelling of in-game events in football. Specifically, we model all touchball events, where a player acts on the ball with some part of their body, as a mutually-exciting point process that depends on the past history of events and occurrence times.

A point process is a probabilistic model for a random collection of points on some space often used to describe the occurrence of random events over time. The Hawkes process is a model for self-exciting point processes, where the chance of a subsequent occurrence is increased for some time period after the initial occurrence. Building on the Hawkes process, our main idea is to take advantage of the decomposition that motivated partial likelihood to separate the modelling of the event types and the occurrence times, thereby overcoming the limitations of traditional point process models. The aim is to provide inferences about previously unquantified measures governing the dynamics of the game as well as predicting the occurrence of events of interest, such as goals, in a specified interval of time.

Keywords

Point processes, Partial Likelihood, Bayesian Inference, Football modelling

Lessons Learned in Scheduling the Finnish Major Ice Hockey League

Kimmo Nurmi, Jari Kyngäs, and Nico Kyngäs

Satakunta University of Applied Sciences cimmo.nurmi@samk.fi jari.kyngas@samk.fi nico.kyngas@samk.fi http://www.samk.fi

Abstract. Ice hockey is the biggest sport in Finland both in terms of revenue and in number of spectators. The Finnish Major Ice Hockey League involves significant investments in players, broadcast rights and merchandising. The quality of the League schedules is extremely important, as the schedule has a direct impact on revenue for all involved parties. We have generated the League schedule for the last ten years. During this time, the League has continuously looked for improvements in its schedule format and the schedule itself. Finding the best schedule of games is a difficult task with multiple decision makers, constraints, and objectives involving logistics, economical and fairness issues. Being able to produce an acceptable schedule is not only about first defining the restrictions, requirements and requests and then developing and using a suitable solution method. An essential part of the problem is the process of consulting with the various league parties. The scheduling process can easily take several months. We believe that scheduling the Finnish Major Ice Hockey League is one of the most difficult sports scheduling problems because it combines break minimization and traveling issues with dozens of other constraints that must be satisfied. We have used the PEAST algorithm and its predecessors to schedule the league since the 2008-2009 season. The algorithm is dynamic in many senses. It is very important that the input of the games need not to be round robins. The algorithm optimizes any number of any games based on the given constraints and goals. For example, in the Finnish Major Ice Hockey League some teams currently play against each other four times and some six times. This paper summarizes the lessons learned in scheduling the League. We present the most important academic and practical findings that we believe will give new ideas to the sports scheduling community.

Keywords

Sports scheduling, PEAST algorithm, Ice hockey

New Performance Measures for the NBA: An Indirect Evolutionary Game Theory Approach

John Obrien and Mark Fichman

Carnegie Mellon University, United States jo0x@andrew.cmu.edu mf4f@cmu.edu

Abstract. Winning games in sport requires making strategic tradeoffs between expected payoffs and risk associated with alternative strategies. Game theory provides a mathematical method of decision-making for analyzing a competitive situation, in order to provide normative insights into such strategic tradeoffs. One branch of this literature, indirect evolutionary game theory (Guth and Yaari (1992), Bisin and Verdier (2001) and Samuelson (2003)), is relevant to competitive sports because preferences are endogenous and winning regulates the evolution of preferences. In professional sports the survival of a coach depends upon their ability to win, which provides a natural testbed for applying the insights from indirect evolutionary game theory to analyzing strategy and developing new performance measures.

In this paper, we use NBA basketball as a testbed for predictions from the indirect evolutionary game theory literature, which provides preference based insights into winning basketball strategies and new performance measures. We first identify the implied Constant Relative Risk Aversion (CRRA) from the closest equilibrium strategy to actual regular season strategies employed during 2016/17 play. We estimate CRRA coefficients from equivalent utility functions for CRRA, using observable mean variance statistics generated from regular season play. Using these estimated preferences for each team we then construct and test the fitness of the implied equilibrium strategies using post season play. Results from this analysis include the following. First, implied CRRA coefficients are risk averse and heterogeneous. Heterogeneity is consistent with theoretical insights from Bisin and Verdier (2001). Second equilibrium strategies predict post-season outcomes, which satisfies our test for fitness. Third, we identify new performance measures, tied to implied preferences that measure a team's defensive strength. A strong defense is able to induce higher levels of risk aversion into the opposing team's offensive strategy. This implies that the opposing team's offence place more weight upon reducing expected payoffs to reduce risk, which lowers their probability of success. In turn, this can also have an interaction effect upon the strong defensive team's offense, which results in reducing their own levels of implied offensive risk aversion to increase expected payoffs and their probability of winning. These results are used to provide a risk preference interpretation of the final playoff series in which GSW defeated CLE in a five game series.

Finally, the risk aversion performance measures are analyzed to address questions such as what are the strategic implications in terms of actual court play, when a strong defensive team is able to induce higher levels of risk aversion from the opposing offensive teams? These types of questions are answered from our decomposition of the court into eleven mutually exclusive zones, by conducting a factor analysis over zones and CRRA coefficients (offensive and defensive). The main factors are then related to win/loss ratios to identify both strong and weak strategies in relation to the CRRA performance measures. That is, we analyze strategy in terms of fitness, preference performance measures and on court zones. These results reveal that there is a subtle interaction between under and overweighting within subsets of both 2- and 3-point court zones.

Keywords

NBA Basketball, Indirect Evolutionary Game Theory, Performance Measures, Risk aversion

Modelling Momentum in Football

Marius Ötting¹, Roland Langrock², and Antonello Maruotti³

- ¹ Bielefeld University, Germany
- ² University of Bielefeld, Germany
- ³ Libera Universita Maria Ss. Assunta, Italy marius.oetting@uni-bielefeld.de roland.langrock@uni-bielefeld.de a.maruotti@lumsa.it

Abstract. Sports commentators and fans frequently use vocabulary such as "momentum", "momentum shift" or related terms to refer to change points in the dynamics of a match. Usage of such terms is typically associated with situations during a match where an event - such as a shot hitting the woodwork in a football match - changes the dynamics of the match, e.g. in a sense that a team which prior to the event had been pinned back in its own half suddenly seems to dominate the match.

Using minute-by-minute in-game statistics of the number of shots on goal and the number of ball touches within Bundesliga matches, we investigate whether such momentum shifts actually do exist in a football match and what kind of events lead to a shift. For that purpose, we consider hidden Markov models (HMMs) for modelling the minute-by-minute bivariate time series data, as they naturally accommodate the idea of a match progressing through different phases, with potentially changing momentum. HMMs involve an unobserved state process modelled by a Markov chain with N possible states and an observed state-dependent process, whose observations are assumed to be generated by one of N distributions according to the Markov chain. Furthermore, within these (multivariate) HMMs, we allow for within-state correlation of our variables by formulating bivariate state-dependent distributions using a copula.

For the investigation of momentum in football, we analyze Bundesliga data from Borussia Dortmund (season 2017/18). Preliminary results suggest a high persistence in the three HMM states considered, with the different states being associated with different playing styles (defensive only, predominantly defensive, and high pressure play). Current research focuses on including covariates in the state process, such that the probabilities of switching between the underlying states depend on (e.g.) the intermediate score of the match and the strength of the opponent.

Keywords

Hidden Markov model, Copula, Football, Momentum

Statistical Models of Horse Racing Outcomes Using R

Alun Owen

Coventry University, United Kingdom aa5845@coventry.ac.uk

Abstract. The published literature on statistical modelling of horse racing outcomes is sparse relative to many other sports, apart from some key seminal works that are now very dated. In addition, there appears to have been little, if any, reference to modelling in horse racing at the MathSport series of conferences since it began as the "IMA International Conference on Mathematics in Sport" in 2007. This lack of published work is not surprising given the potential commercial value of such knowledge and hence the potential unwillingness for authors to share their work.

The aim of the current work is therefore to offer an overview of the current literature in the area of statistical modelling in horse racing outcomes, and to present details of how a potentially profitable model can be developed using a discrete choice modelling approach within the statistical computing environment R (R Core Team, 2013). This will include considerations of suitable data sources and model development and implementation in R, as well assessments of model fit and potential profitability of the model when used for betting. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Keywords

Statistical Modelling, Discrete Choice Modelling, Horse Racing, Betting,

Bayesian Modelling of Volleyball Sets

Vasileios Palaskas¹, Ioannis Ntzoufras¹, and Sotiris Drikos²

- ¹ Department of Statistics, Athens University of Economics and Business, Greece
- ² Faculty of Physical Education and Sport Sciences, National and Kapodistrian University of Athens, Greece bill_pala@hotmail.com

ntzoufras@aueb.gr sdrikos@gmail.com http://www.stat-athens.aueb.gr/~jbn https://www.sdrikos.com/

Abstract. In this work, we focus on building a Bayesian model to analyze the outcome of a Volleyball game as recorded by the difference in winning sets. More specifically, the first and foremost challenge is to find the appropriate model for the response outcome (sets) which cannot be based on the usual Poisson or Binomial assumptions. Here we will use two major approaches: a) a model based on a truncated version of the Skellam Distribution b) an Ordinal Multinomial logistic regression model. Concerning the first model, we adjust the Skellam distribution in order to take into account for the Volleyball rules. More specifically, in the light of the non-existence of ties and under the fact that the winning team has maximum three sets in Volleyball games, we need to adjust this model in order to deal with these issues. For the second model, we will consider the set difference as an ordinal response variable within the framework of Multinomial logistic regression models. As far as covariates we consider, at first we will fit and compare the models with a common home effect, and the usual attacking and defending team abilities as implemented in other predictive models for sports (see for example in Karlis and Ntzoufras, 2003). In the second level, we will consider additional covariates that capture the characteristics of the game such as the serve, the block, the attack of type I and type II, etc. Bayesian model comparison will be implemented by WAIC and/or DIC. Our aim is also to incorporate modern Bayesian variable selection algorithms and by this way also implement automatically Bayesian model averaging. The data set in our case concerns the Greek A1 professional men's League of the regular season 2016-2017 recorded and analyzed through software by one observer registering every touch of the ball during the game. For each set in the entire League (N=494) we record the data for both teams. The main software tools we will use are the statistical software of R and the probabilistic programming language of Stan which is proper for fitting Bayesian models. Furthermore, WINBUGS may be used alternatively to model specific characteristics of the problem.

Keywords

Bayesian Modelling, Volleyball, Skellam and Multinomial regression models

Quantifying the Evolution of First-Class Rugby in New Zealand

Ankit Patel, Paul Bracewell, and Jack McIvor

DOT Loves Data, New Zealand ankit@dotlovesdata.com jack@dotlovesdata.com

Abstract. Rugby in New Zealand has evolved since the first recognised game was played within the country on 14 May 1870. This is observed within the domestic first-class rugby which has been influenced by the introduction of the National Provincial Championship in 1976, the subsequent format changes of that competition as well as rugby coming professional in 1995. To quantify this change, Moore et. al., (2018) approach to optimally defined the ratings update parameter, k, within a modified Elo rating framework is expanded to dynamically adjust the underlying update parameter, k, within a competition. This is further enhanced by applying the geometric distribution to perform bootstrapping on the time series and create confidence intervals for k. Results from first class games between New Zealand's 27 provinces between 1978 and 2018 are used. The calculated ratings and match prediction are then contrasted against Glicko2 to validate the applicability of this approach. Finally, the changes in k are interpreted to provide further insight about the underlying mechanics of domestic first-class rugby within New Zealand.

Keywords

Elo ratings, Glicko2, Rugby, Cross-validation, optimization

Detection of Playing Style of Football Teams and Players using Latent Dirichlet Allocation

David Perdomo Meza

Football Whispers, United Kingdom dperdomomeza@gmail.com

Abstract. This paper introduces an innovative unsupervised method based on Latent Dirichlet Allocation (LDA) to learn and then detect football team and player playing styles. LDA is borrowed from the subject of Topic Extraction in Natural Language Processing where it is implemented to automatically sort a corpus of text documents into a mixture model of the semantic topics that constitute them. Once the model is fit, the trained model can be applied to new documents it had not seen before and sort them into constituting topics. In our approach, we repurpose this model and train it on aggregate match data to sort teams/players into a mixture model of different playing styles. As in the NLP implementation the methodology is unsupervised, and yet the results have proved to be closely aligned with natural concepts of playing style in football. The fit model has proven itself to be a fascinating, scalable tool to provide stylisitic information on teams and players.

Keywords

Playing Style, Topic Models, Natural Language Processing, Latent Dirichlet Allocation

Winning in Basketball with Data and Machine Learning

Konstantinos Pelechrinis¹

University of Pittsburgh, United States kpele@pitt.edu http://www.sis.pitt.edu/~kpele

Abstract. 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 models 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.

Keywords

Expected possession value, Basketball analytics

Multi-Criteria Solutions for Optimizing Lineup in Baseball

Marina Polyashuk

Northeastern Illinois University, United States m-polyashuk@neiu.edu

Abstract. This paper focuses on the task of creating an optimal lineup for a baseball game, which is one of the most important and consequential tasks facing a Major League Baseball team manager. The lineup problem is decomposed into two parts: first, selection of an optimal 9-player set of defensive position players, and second, choosing the best way to arrange this set of nine players in the batting order. The problem of choosing a set of starter players for a game is formulated as a bi-criteria problem with such criteria as the offensive potential of the set and the defensive quality of the set, with optimal solutions defined as non-dominated (Pareto-optimal) solutions with respect to both criteria. The problem of the batting order is viewed as a combinatorial problem with two lexicographic-ordered criteria, the first of them being the expected number of runs scored by one rotation of the lineup. Throughout the paper, the proposed methodology is applied to the example of the 2012 Chicago White Sox team; actual managing decisions are compared to the solutions proposed by our model.

Keywords

Multi-criteria solutions, Optimal Lineup in Baseball, Optimal starting player selection, Optimal batting order, Pareto-optimal solutions, Expected number of runs scored

Perspectives on Fairness in Sport

Chris Potts

University of Southampton, United Kingdom C.N.Potts@soton.ac.uk http://www.soton.ac.uk/maths/people/profiles/or/cnp1.html

Abstract. In a match that forms part of a high-profile sports tournament or competition, there are often several stakeholders who may hold conflicting views about some aspects of the match. The stakeholders include the players who are competing, the spectators, the TV company that may invest a substantial sum of money to broadcast the match, and the governing bodies that run the sport.

First, we review two studies in the sport of golf. One study focuses on the cut rule for the PGA that determines which players "make the cut" after the first two rounds of the tournament and consequently proceed to play the final two rounds. In terms of player fairness, players with a realistic chance of a high finishing position should not be eliminated, whereas a major concern of TV companies is that the number of players should not be so large that there is a high likelihood that on the final day the tournament does not finish within the assigned TV slot. The other study proposes an alternative match-play format for the FedEx Cup, the end-of-season playoff event on the PGA tour. The current system (although it is to be changed in 2019) is based on points gained for performance during both the regular season and the playoffs. However, it seems impossible to design a points allocation that adequately rewards consistent play throughout the season and also gives all players competing in the final tournament a chance of winning. Therefore, the system is unfair to the spectators who may well not see an exciting end to the season, and it is dissatisfying for those players who cannot become the overall winner. The match-play system overcomes these disadvantages.

The talk finishes with some examples taken from various sports where unsatisfactory rules for the competition have resulted in unfortunate consequences for the players and/or spectators. Also, some successes are highlighted, where changes of rules or format have produced a better experience for both the competitors and spectators.

Keywords

Fairness, Multiple stakeholders, Format/rule changes

Finding optimal strategies for substitutions in soccer using a two-scale Markov Decision Process approach

Jörg Rambau and Rónán Richter

Lehrstuhl Wirtschaftsmathematik, Universität Bayreuth Joerg.Rambau@uni-bayreuth.de Ronan.Richter@uni-bayreuth.de

Abstract. While a coach in soccer has much influence on his team's chances of winning before the match, e.g. by designing the training, giving tactical plans or choosing the starting line-up, his abilities during the match are very limited. One of his main ways of influence are strategic substitutions. A substitution will be called strategic, if it aims to change the formation of the team. Possible questions in that context are: In which cases should a coach send an additional defensive player onto the field in order to save a lead or when to introduce a striker for more offensive possibilities? Questions of this type can be tackled by modeling a soccer match as a twoscale Markov Decision Process (2MDP), that consists of a detailed MDP modeling the gameplay and a coarse MDP modeling the strategic decisions. This concept has recently been introduced by Hoffmeister and Rambau, who used beach volleyball as a first application. One main feature of this model is, that it does not relay on historic data on the success of the different decision rules against a specific opponent, that is rarely available, but only on data of the players' abilities, that can be observed independent of other influences. In this talk, a 2MDP model for strategic substitutions in soccer is presented. The differences to net sports are outlined. Additionally, first results for optimal substitution strategies on artificial data are given and the importance of decisions on substitutions in various scenarios is discussed.

Keywords

OR in Sports, Markov Decision Process, Dynamic Programming, Decision Support Systems, Soccer Substitutions

Anti-Analytics in Football

Anay Rennie¹ and Jayanth Kolla²

¹ Deloitte,	India
------------------------	-------

² Convergence Catalyst, India anayrennie@gmail.com anayrennie@gmail.com http://convergencecatalyst.com/

Abstract. Football has become increasingly data-driven and these days almost every team uses rigorous analytical techniques to enhance their own performances, strategize on opponents performances and consequently increase their winning chances. We have structured the whole process flow with deep-dive research and further went on to build an anti-analytics technique to prevent other teams from using analytics on oneself. This will help a team prevent their gameplay from being analyzed completely and consequently have an upper hand on the analytics game compared to their oppositions.

Keywords

Anti-Analytics, Comparative Analytics, Football, Co-Ordinate Geometry, Strategy

Iterated Local Search for Multi-league Scheduling Problems

Panagiotis Repoussis¹ and Walter Gisler²

- ¹ Athens University of Economics and Business, Greece
- ² GotSoccer LLC, United States prepousi@aueb.gr walter@gotpro.com

Abstract. Tournament and league scheduling problems are notoriously hard to solve. They may involve a wide variety of ordering and assignment constraints with conflicting requirements (e.g. fairness, police requests, minimize travelling, competitiveness and availability of venues). For this reason, it is often hard to find feasible solutions that satisfy all constraints, and thereafter, the goal is typically to minimize the penalties that occur from soft constraint violations. This work focuses on multi-league scheduling problems with cross-league pairing constraints (e.g. specific matchups or number of games that can be played at a given location on a specific period considering two or more leagues). An Iterated Local Search (ILS) is proposed that employs a sophisticated ruin-and-recreate perturbation mechanism. The inner local search framework is equipped with neighborhood structures based on intra- and inter- period exchanges, while short- and long-term adaptive memory structures are also employed to guide the search process. Preliminary results on realistic problem instances are promising, demonstrating the efficiency and effectiveness of the proposed approach.

Keywords

League Scheduling, Tournament Scheduling, Tabu Search, Professional Sports, Iterated Local Search, Heuristics

A Bayesian dynamical bivariate Poisson state space model for predicting football scores and results

Peter Ridall

Lancaster University, United Kingdom g.ridall@lancs.ac.uk

Abstract. We derive fast updates for a Bayesian dynamical bivariate Poisson state space model for modelling football scores and outcomes. This has similarities to the dynamic GLMs of Harrison and West but uses different approximations. Exploiting conjugacy, we use the expectation of the full conditional posteriors to construct upfdates for the sufficient statistics for the dynamic parameters. We introduce fixed forgetting parameters to describe the volatility of the form of the sides both between the season and within the seasom. These parameters can be found quickly by maximisng the evidence. In this way we are able to track the form and style of each of the clubs over the last two decades identifying periods of rapid and minimal change.

Keywords

State space models, Sequential Bayesian methods, Conjugacy

Modelling netball scores

Phil Scarf¹, Rose Baker¹, and Rishikesh Parma²

- ¹ University of Salford, United Kingdom
- ² IIT Dhanbad, India
 - p.a.scarf@salford.ac.uk
 - r.d.baker@salford.ac.uk
 - rishikesh786@am.ism.ac.in

Abstract. We model the scores in Netball matches. Our model borrows some parts from the Poisson match-the notion of attacking and defensive strengths-and some parts are new-three sources of randomness: the total score; and the conversion of possession into goals by each team. In this way, the structure of netball leads to what we think is an interesting model. We estimate model parameters using maximum likelihood and for a UK tournament, the Vitality Superleague, which has been in-play for about a decade. Match data are somewhat sparse in parts, as new teams have entered the league and others have left. Therefore, our estimation method has an element of shrinkage. Strength varies over time so we use a discounted likelihood. With this model, we consider questions related to outcome uncertainty, tournament design, tactical play, and forecasting.

Keywords

netball, poisson match, outcome uncertainty
Predicting possession outcomes using spatiotemporal data in Australian Rules football

Bartholomew Spencer¹, Karl Jackson², and Sam Robertson¹

¹ Victoria University, Australia

² Champion Data, Australia bartholomew.spencer@vu.edu.au karl.jackson@championdata.com.au sam.robertson@vu.edu.au

Abstract. Introduction: The expected outcome of possessions has been a commonly researched topic in various sports. In this study, we develop and compare models for computing the expected result of possession chains in Australian Rules football. For each gain of possession, the possession expectation is computed based on spatial characteristics of both the current and preceding possessions. A secondary interest was in identifying changes in possession expectation as a result of changes in density and spatial control at the preceding possession - that is, can drawing opponents in (hence, reducing immediate space) create future valuable space? Methods: LPS player tracking and play-by-play transactional datasets were consolidated for 40 matches in the 2018 AFL season. Spatial characteristics of the attacking team and their opponent were computed for each possession. These included measures of spatial control (e.g., surface area, centroids, team densities and subsequent dominance), team displacements (e.g., velocity, angular displacement, orientation, phase coupling), and positional information. Linear and non-linear predictive models were compared on inputs of varying complexity. A Random Forest model was run to examine variable importance. Results: Models were fit on 15397 possessions (60/20/20 split for)training/validation/testing datasets). Non-linear approaches produced lower RMSE than linear ones. Additionally, models fit on inputs containing measures of spatial control outperformed those fit only on positional information. Team displacements in the preceding one second period, opposition orientations, and team centroid locations had the highest variable importance. Changes in density during the preceding possession were found to have low relative importance. Conclusions: The addition of spatiotemporal data improves the predictive power of possession expectation models in Australian Rules football.

Keywords

player-tracking, spatiotemporal, AFL, predictive modelling

Ranking Elite Swimmers using Extreme Value Theory

Harry Spearing and Jonathan Tawn

Lancaster University, United Kingdom
h.spearing@lancaster.ac.uk
j.tawn@lancaster.ac.uk
http://www.lancaster.ac.uk/ spearing/

Abstract. The International Swimming Federation (FINA) uses a very simple points system with the aim to rank swimmers across all Olympic events. The points acquired is a function of the difference between the recorded time and the current world record for that event. With some world records considered "better" than others however, bias is introduced between strokes, with some being much harder to attain points where the world record is hard to beat. I will introduce a model based on extreme value theory, where the distribution of swim-times is viewed as a Poisson point process, with the distribution of the best times following a Generalised Pareto distribution. Within this framework, the strength of a particular swim is judged based on its position compared to the whole distribution of swim-times, rather than just the world record. This model also accounts for the date of the swim, as training methods improve over the years, as well as changes in technology, such as swim-suits. The parameters of the Generalised Pareto distribution, for each of the 34 individual Olympic events, are pooled into a hierarchical model structure, by exploiting patterns found when introducing covariates into the model. This allows information to be shared across all strokes, distances, and genders, so that predictive power is improved, and model fit is more robust. A by-product of the model is that it is possible to estimate other features of interest, such as the ultimate possible time and the distribution of new world records for any event. I will illustrate the methods using the top 200 best times for each event in the period 2001-2019.

Keywords

Extreme Value Theory, Smoothing Splines, Statistical Modelling, Swimming

The Ancient Olympics: Events, Technology, Superstars, Women, Lessons for Them and for Us

Raymond Stefani

California State University, Long Beach, United States raymond.stefani@csulb.edu

Abstract. A multidimensional analysis of the ancient Olympics is presented. Originating in 776BC, the Olympic Games have emerged as the best known and documented of the four Panhellenic Games, compared to the Nemean, Isthmian and Pythian Games. A reliable list of 861 Olympic events and winners was analyzed. Among athletics events, were the stadion (a run of about 200m), the diaulos (about 400m), the diaulos in armor and the pentathlon (stadion, discus, javelin, long jump and wresting). There were combat events (boxing, wrestling and the no-holds-barred pankration), chariot racing, equestrian racing and artistic performances (herald competition, trumpeting, lyre playing and acting). Significant technology included the clever use of ropes and levers to start the running events and the chariot races. A cord was wrapped around the javelin with a finger loop to create spin stabilization. Beginning with a standing start, long jumpers employed complex kinematics to extend their distances while carrying weights. The three greatest superstars were Leonides of Rhodes who won all three of the main running events four times in a row (12 wins), Herodoros of Megara who won the trumpeter competition nine times in a row and Astylos of Croton who won 7 athletics events. Five athletes won six times, including Nero (whose wins might have been somewhat contrived). Those superstars would have much to teach us as to training methods and techniques, while our video analysis and knowledge of nutrition could have helped them. Although the Olympic Games were only open to men, Kykniska of Sparta, a married woman, was a double Olympic champion, having twice owned and trained winning chariot horses. Women competed at Olympia in their own separate Heraean Games. The running distances were shortened from multiples of 600 Greek feet (for men) to multiples of 500 Greek feet (for women). In today's world, the Olympic flame is lit at Hera's shrine, providing women with a magnificent symbol of equality.

Keywords

Ancient Olympics, athletics, combat sports, women in sports, kinesiology, technology, long jump, pentathlon, pankration, trumpeting

Raking of Japanese University Baseball

Takashi Toriumi

Keio University, Japan toriumi.takashi@gmail.com

Abstract. The Japan University Baseball Federation consists of 26 baseball leagues in all over Japan. Most leagues have some divisions and the most top ones consist of six teams; however, there is no opportunity for the teams in these leagues to play directly against the teams belonged to the other leagues except only the matches of the two national championships, Japan National Collegiate Baseball Championship (NCBC) and Meiji Shrine Baseball Championship (MSBC), and those of their preliminary rounds. It is important for the fans and high school baseball players to rate and rank the university baseball teams across the leagues. In this report, we calculate the ratings for each team belongs to the 26 leagues. We use data of matches conducted within the 2192 matches of the 26 top leagues in 2018 season. Using the Elo rating calculate, we estimate the ratings of 187 universities in each league as follows: 1) We calculate the ratings of six teams in the Tokyo Big6 Baseball League that is one of the most popular and competitive leagues among the 26 leagues with the various k values and decide the most appropriate k value with the highest probability that the higher rating teams win; 2) We estimate the strengths of each team based on the results in each league with the k value. The results of step 1 indicate that the most appropriate k value is 38 and the probability is 63.9%. The results of step 2 indicate that the team with the highest rating in the spring season, first half of the season, is the champion of the national championship (NCBC) however, the team with the highest rating in the autumn season is not the champion of the national championship (MSBC). That is because the ratings of the autumn season include the results of spring season. We found that the winning rates, the number of wins and win or lose in the latter of the season affected a lot the final ratings.

Keywords

Japanese Baseball, University, Elo rating, ranking

Objectively Modelling the College Football Playoff Committee's Selections

John Trono

Saint Michael's College, United States jtrono@smcvt.edu http://academics.smcvt.edu/jtrono

Abstract. At the end of the NCAA college football season, four teams are selected (by this committee) so that those teams can then compete in the single elimination, College Football Playoff (CFP) which crowns that sport's national champion. Thirty two models have been evaluated over the first five years of the recently created CFP, with regards to matching how the CFP committee has ranked its top 25 teams by varying the victory reward parameter (four choices) and the group-size parameter (eight choices) when employing one possible strategy to assess each team's success that year. Those 32 models only consider wins and losses, whereas a slightly modified set of 32 models includes margin of victory (MOV) as well. One of those models-both with, and without, MOV-has correctly predicted (the same) 18 of the 20, top four teams that the CFP committee has chosen the past five years, and another eight models (five using MOV) have matched 17 out of 20. These various models, and a deeper analysis of them, will be presented here.

Keywords

Parameterized Modelling, Team Ranking, Matching human experts

Combining the Four Factors with the Generalized PageRank (GeM) model for NBA Basketball

Georgia Twersky, George Lyman, and Anne Pilkington

University of Notre Dame, United States gtwersky@nd.edu glyman@nd.edu pilkington.4@nd.edu

Abstract. In their 2008 paper on the use of a generalization of Google's page rank algorithm to rank teams in the NFL, Meyer et. al. suggest using the algorithm with a variety of game statistics and pose several questions at the end of the paper as to which statistics should be included, and how the resulting stochastic matrices should be weighted in the model. In this study, we used the generalized version of Meyer's Google's PageRank (GeM) ranking system, combined with Oliver's Four Factors for basketball to make predictions for outcomes in the NBA playoffs. In particular, we compared the performance of this ranking system with the performance of GeM when calculated with the point differential and with the performance of Massey's ranking system with the point differential. We also created a ranking using Massey's system and the Four Factors for comparison. We ran all models on the NBA playoff data from 2007 to 2018. The results show no significant difference in the performance of the models using the four factors and the models using the point differential.

Keywords

ranking, performance models, basketball

Using the Poisson-Lognormal Regression Model with Varying Dispersion to analyze soccer data

George Tzougas¹ and Dimitris Karlis²

 $^{1}\,$ London School of Economics and Political Science, United Kingdom

² Athens University of Economics and Business, Greece g.tzougas@lse.ac.uk karlis@aueb.gr http://www.lse.ac.uk/Statistics/People/Dr-George-Tzougas http://www2.stat-athens.aueb.gr/~karlis/

Abstract. The literature contains several count models for the number of goals in a soccer match. In this paper we consider the Poisson-Lognormal regression model with varying dispersion for the distribution of the number of goals scored in the English Premier League. The new feature of the model is that we assume that the overdispersion parameter is affected by covariate information and hence each team can have different overdispersion. This adds flexibility to the model. Our main contribution is that we develop a simple Expectation-Maximization algorithm for maximum likelihood estimation of the model when regression specifications are used for every parameter. Application to real data from Premiership under the proposed modelling framework is considered. Finally, extensions of the model are discussed.

Keywords

Poisson-lognormal Regression Model, Varying Dispersion, EM Algorithm, Soccer

Sports betting strategies: an experimental review

Matej Uhrín, Gustav Šourek, Ondřej Hubáček, and Filip Železný

Czech Technical University in Prague uhrinmat@fel.cvut.cz souregus@fel.cvut.cz hubacon2@fel.cvut.cz zelezny@fel.cvut.cz http://ida.felk.cvut.cz/zelezny

Abstract. We investigate the problem of optimal wealth allocation over predictive sports market's opportunities. We analyze the problem across diverse input variables, target utility measures, and the notion of optimality itself. We review existing literature to identify the most prominent approaches coming from the diverse sport and economic views on the problem, and provide some practical perspectives. Namely, we focus on the provably optimal geometric mean policy, typically referred to as the Kelly criterion, and Modern Portfolio Theory based approaches leveraging utility theory. From the joint perspective of decision theory, we discuss their unique properties, assumptions, and computational complexity and, importantly, investigate effective heuristics and practical techniques to tackle their key common challenges, particularly the problem of uncertainty in the outcome probability estimates, for which we experimentally review methods such as distributionally robust convex programming [1], fractional Kelly [2], and maximum drawdown constraints [3]. Finally, we verify our findings on simulated datasets as well as real data coming from three different sports betting domains.

[1] Sun, Qingyun, and Stephen Boyd. "Distributional Robust Kelly Gambling." arXiv preprint arXiv:1812.10371 (2018).

[2] MacLean, L. C., William T. Ziemba, and George Blazenko. "Growth versus security in dynamic investment analysis." Management Science 38.11 (1992): 1562-1585.

[3] Busseti, Enzo, Ernest K. Ryu, and Stephen Boyd. "Risk-constrained Kelly gambling." arXiv preprint arXiv:1603.06183 (2016).

Keywords

sports analytics, betting markets, betting strategies, uncertainty, geometric mean criterion

Identifying Technical Functions of Footballers using Hierarchical Cluster Analysis

Orbay Unsoy

The University of Manchester, United Kingdom orbayunsoy@gmail.com

Abstract. In football, conventional playing positions provide a rough idea on the tasks players are expected to perform. Yet, the evolving characteristics of the game call for an enhanced understanding of player contributions. This study sets out to identify technical functions of footballers based on their skill sets.

Research methods: The technical performance indicators of 920 players fielded in the top three European football leagues during the 2017/18 season were analyzed via principal component analysis and agglomerative hierarchical clustering.

Results and findings: In the principal component analysis phase, we introduce three off the ball and five on the ball skill sets to characterize the contributions of footballers. Later, we identify seven major clusters and 18 sub-groups that can be utilized to categorize technical player functions. We then illustrate how the model could be beneficial in supporting the recruitment decision making of football players.

Implications: The findings imply that a function-based categorization can help to overcome systematic biases caused by heuristics based on conventional positions and physical attributes of players. For key decision makers in football clubs, our model provides a common language on player functions based on empirical evidence. Further, the derived skill sets allow to assess talent comprehensively within the context of the analyzed competitions.

Keywords

cluster analysis, football analytics, player functions, player recruitment, talent management

A hybrid random forest to predict soccer matches in international tournaments

Hans Van Eetvelde¹, Christophe Ley¹, Andreas Groll², and Gunther Schauberger³

² Technische Universität Dortmund, Germany

³ Technical University of Munich, Germany hans.vaneetvelde@ugent.be christophe.ley@ugent.be groll@statistik.tu-dortmund.de gunther.schauberger@tum.de

Abstract. We propose a new hybrid modeling approach for the scores of international soccer matches which combines random forests with Poisson ranking methods. While the random forest is based on the competing teams' covariate information, the latter method estimates ability parameters on historical match data that adequately reflect the current strength of the teams. We compare the new hybrid random forest model to its separate building blocks as well as to conventional Poisson regression models with regard to their predictive performance on all matches from the four FIFA World Cups 2002-2014. It turns out that by combining the random forest with the team ability parameters from the ranking methods as an additional covariate the predictive power can be improved substantially. Finally, the hybrid random forest is used (in advance of the tournament) to predict the FIFA World Cup 2018. To complete our analysis on the previous World Cup data, the corresponding 64 matches serve as an independent validation data set and we are able to confirm the compelling predictive potential of the hybrid random forest which clearly outperforms all other methods including the betting odds.

Keywords

FIFA World Cup 2018, soccer, random forests, prediction, team abilities, sports tournaments

¹ Ghent University, Belgium

Pre-Game Speeches: Impact and Measurement

Tiffanye Vargas

Long Beach State University, United States Tiffanye.Vargas@csulb.edu

Abstract. Coaching behaviors, intentionally and/or unintentionally, have a strong impact on athletic performance. This can result from the coaching climate they foster, the relationships they hold with their athletes, or the culture they create within their team. Additionally coaching behaviours may influence changes in athletes' perceptions of self- efficacy. Self-efficacy is a person's belief in his/her ability to complete a specific task (Bandura, 1977; 1997). Sources of efficacy include performance accomplishments, vicarious experiences, verbal persuasion, and physiological arousal. For coaches, the source of verbal persuasion is a readily available tool that can be used to impact athletes' efficacy, and resultant behaviours. A coach's use of pre-game speech, a form of verbal persuasion, has the potential to be a secret weapon in readying teams for performance; however, it has just as much potential to help teams self-destruct. Previous research has found that informational content in pre-game speeches is important (Vargas & Short, 2011), as is emotion in the delivery of such speeches (Vargas & Bartholomew, 2006). Athletes believe pre-game speeches to be effective (Vargas & Guan, 2007; Vargas & Short, 2011), and athletes' prefer different types of pre-game speech content (i.e., emotional vs information) according to the situation (Vargas & Guan, 2007). However, much remains unknown about the pre-game speech and its effectiveness. Therefore, the purpose of this presentation will address the current research on this particular form of verbal persuasion. Specific attention will focus on the difficulties of measuring the effectiveness of the pre-game speech and its resultant effect on athlete efficacy. This is one of four presentations addressing sport psychology and measurement.

Keywords

Sport psychology, Coach behavior, pre-game speech, self-efficacy

Avoiding combinatorial clashes for the Champions Hockey League Group Stage Draw

Stephan Westphal¹ and Martin Dahmen²

¹ Clausthal University of Technology, Germany ² TU Clausthal, Germany stephan.westphal@tu-clausthal.de martin.dahmen@tu-clausthal.de http://www.mathematik.tu-clausthal.de/mitarbeiter/stephan-westphal/

Abstract. For the season 2016/2017, the Champions Hockey League (CHL) was performing a Group Stage Draw in which 48 teams had to be drawn out into 16 groups of three teams each. The allocation of the teams to the groups had to be done in a way such that teams coming from the same league were not drawn out into the same group. Furthermore, clubs from leagues where the national teams were participating in the final olympic qualification tournament could also not be drawn out into the same group.

At first, the 16 teams from pot 1 were drawn out into the groups A-P such that the first team drawn is assigned to group A, the second one to group B and so on. Then, the 16 teams from pot 2 were being drawn out in the same way, followed by the 16 teams from pot 3. But, whenever such an assignment would have led to two teams from the same league playing in the same group, the current group was skipped and the team was allocated to the next group in which it did not cause such a "league conflict". The same applied for assignment leading to "olympic qualification conflicts".

Whenever the assignment of a team to a group is not possible because it raises a league conflict or an olympic qualification conflict in this particular group, we will call this a "direct conflict". These are easy to see and can thus be prevented easily. But there are also "combinatorial conflicts" which are hard to detect. These are conflicts which arise when a team is allocated to a group in which it does not provoke any direct conflict but the set of remaining teams in the pots cannot be assigned without direct conflicts anymore.

Since the Draw was broadcasted live, the CHL needed to know immediately if an assignment would have led to a direct or a combinatorial conflict and to which group the teams had to be allocated directly after they have been drawn out of the pots.

In this talk we discuss the algorithm based on an integer programming model, which was used by the CHL for the 2016/17 Group Stage Draw.

Keywords

decision support systems, mathematical and physical models in sports, optimal tournament design and scheduling

Contextual information improves the accuracy of predicting the direction of serve in professional tennis

Hiroyuki Yamamoto¹, Kazutoshi Kudo¹, Tim Buszard², Machar Reid², Damian Farrow², and Stephanie Kovalchik²

- ¹ The University of Tokyo, Japan
- ² Institute of Sport, Exercise and Active Living, Victoria University, Melbourne, Australia hyamamoto.mcml@gmail.com

Abstract. The purpose of this study was to examine whether Contextual information (CI) improved prediction accuracy of serve direction (to receiver's forehand (FH) or backhand (BH)). Several categories of CI variables were considered: player characteristics, player performance, score, environment and prior serve directions. The data sample included 999 matches and 75538 points from male professional singles matches from 2013 to 2017. The servers in the sample included 25 players ranked in the top 30. Results show that left handed servers target the BH more than right handed servers (p < .05), and there is a significant negative correlation between average aces in a match and BH proportions (r = -0.57, p < .01). A prediction model for serve direction was built using a Markov model whose states were the backhand or forehand direction serve. Our prediction model increased prediction accuracy for the serve direction of five players comparing to naive model without CI. CI can provide strategic advantage for predicting the service behavior of elite opponents.

Keywords

FIFA World Cup 2018, soccer, random forests, prediction, team abilities, sports tournaments

Structuring Patterns in Team Movement and Performance from Vector Field Analysis and Fractal Dynamics

Abdullah Zafar and Farzad Yousefian

Sport Performance Analytics Inc., Canada zafar.abdullah10@gmail.com fyousefi3@gmail.com

Abstract. Team movement analysis in football is critical in evaluating performance. A vector field approach to model team movement dynamics can provide insight into the overall team state by considering the spatio-temporal information of individual players. The aim of this study was to understand the relationship between team movement dynamics and player physical performance measures using vector field analysis.

Female university football players (n=17) were monitored during five competitive matches at an international tournament using combined 10-Hz global positioning system (GPS) and Heart Rate recordings (Polar Team Pro System, Polar Electro Oy, Finland). A radial basis function of the coordinates of 10 outfield players generated a dense vector field. The gradient of the field divergence determined the intensity of team displacement, termed the driving force (DF). Team "tempo"t was quantified using a continuous wavelet transform (CWT) on the DF time series. A Detrended Fluctuation Analysis (DFA) was applied to the DF time series, and DF was correlated to physical performance metrics for 15-minute periods of the match using Pearson correlation coefficients.

DFA showed an average Hurst exponent of H 0.62 and 0.54 in the first and second half respectively and an average persistence time of 32 seconds (18-50 seconds). Tempo dropped an average of 6% (1-12%) between halves. Significant (P < 0.05) positive correlations were observed between DF and work-rate (m/min) (r = 0.83, 95% CI 0.63 - 0.92) and high-intensity running ($\geq 16.0 km/h$)(r = 0.71, 95% CI 0.49 - 0.84). A zonal analysis of force, tempo and the multifractal spectrum of the DF revealed specific positional dependencies in team movement dynamics.

These results suggest the presence of Fractional Brownian motion in football team movement dynamics; and such behaviour is related to specific physical performance measures. Practitioners in football can use this information in implementing tactical and physical training protocols.

Keywords

Football, Performance Analysis, Team Movement Dynamics

Index

Ötting, M., 49 Šourek, G., 24, 69 Železný, F., 24, 69 Alexander, J., 1 Atan, T., 2, 7 Baker, R., 61 Barbiero, A., 3 Bellio, R., 20 Besson, R., 4 Beyer, R., 5 Bleak, C., 13 Blevins, A., 26 Bracewell, P., 52 Brooks, R., 8 Buszard, T., 74 Caneco, B., 13 Cant, O., 6 Carlsson, N., 17 Cavdaroglu, B., 2, 7 Collingwood, J., 8 Cross, R., 6 Csato, L., 9 Cueva, V., 10 Cunha, S.A., 44 Curley, B., 11 Dahmen, M., 73 Danzig, L., 12 Dellaportas, P., 46 Demir, E., 21 Donovan, C., 13 Drikos, S., 14, 51 Egidi, L., 15 Farrow, D., 74 Fichman, M., 48

Fonseca, G., 16, 20 Günneç, D., 21 Gisler, W., 59 Giummolé, F., 16 Goes, F., 18, 30 Goosens, D., 19 Goossens, D., 37 Grassetti, L., 20 Groll, A., 18, 71 Guyon, J., 22 Hirotsu, N., 23 Hopkirk, G., 11 Hubáček, O., 24 Hubáček, O., 69 Hunt, M., 26 Hunter, G., 25 Huynh, M., 29, 45 Hvattum, L.M., 43 Ingram, M., 27 Jackson, K., 62 Jauvion, B., 25 Jenkins, S., 28 Jonker, L., 38 Karlis, D., 39, 68 Kee, L., 29 Kempe, M., 18, 30 Koevoets, W., 31 Kolla, J., 58 Komine, A., 23 Konaka, E., 32 Kosmidis, I., 33, 46 Kouřim, T., 34 Kovalchik, S., 6, 35, 74 Kudo, K., 74 Kyngäs, J., 47

Kyngäs, N., 47 Lambers, R., 36, 37 Lambrix, P., 17 Langrock, R., 49 Lawrence, S., 38 Lemmink, K., 18 Leriou, I., 39 Leroy, A., 40 Ley, C., 71 Lokhorst, R., 11 Lyman, G., 67 Madrigal, L., 42 Mara, J., 1 Maras, K., 41 Marshall, K., 29, 45 Maruotti, A., 49 McIvor, J., 52 Merlin, M., 44 Meyer, D., 29, 45 Meza, D.P., 53 Moura, F., 44 Narayanan, S., 46 Ntzoufras, I., 15, 39, 51 Nurmi, K., 47 Obrien, J., 48 Olmo-Jiménez, M.J., 10 Owen, A., 50 Palaskas, V., 51 Pantuso, G., 43 Parma, R., 61 Patel, A., 52 Pelechrinis, K., 54 Pilkington, A., 11, 67 Poli, R., 4 Pollard, G., 45 Polyashuk, M., 55 Potts, C., 56 Principe, V., 44

Rambau , J., 57

Rashid, I., 25 Ravenel, L., 4 Reid, M., 6, 74 Rennie, A., 58 Repoussis, P., 59 Richter, R., 57 Ridall, P., 60 Robertson, S., 1, 62 Rodríguez-Avi, J., 10 Sans Fuentes, C., 17 Scarf, P., 61 Schauberger, G., 17, 38, 62, 71 Sharif-Ali, M., 25 Spearing, H., 63 Spencer, B., 1, 62 Spieksma, F., 36, 37 Stefani, R., 64 Sweeting, A., 1 Tawn, J., 63 Toriumi, T., 65 Torres, R., 44 Trono, J., 66 Twersky, G., 67 Tzougas, G., 68 Uhrín, M., 69 Unsoy, O., 70 Van Eetvelde, H., 71 Vandebroek, M., 19 Vargas, T., 72 Verbeek, J., 38 Vidoni, P., 20 Wang, C, 19 Westphal, S., 73 Wright, M., 8 Yamamoto, H., 74 Yousefian, F., 75 Zafar, A., 75

The photo in the preamble is a courtesy of Vaggelis Anatsasopoulos https://www.instagram.com/anastasopoulosvaggelis/