10th MathSport International Conference Proceedings 2023

10th MathSport International Conference Proceedings 2023

Corvinus University of Budapest

Budapest, 2023

10th MathSport International Conference Proceedings 2023

Editor: László Csató

Authors:
Ethan Baron, Bhaskar Basu, Yingjiu Bei, Matthias Bogaert, Jimut Bahan Chakrabarty, Kerem Demirtaş, Ning Ding, László Edvy, Keisuke Fujii, Marc Garnica Caparrós, Jim Griffin, László Gyarmati, Laurentiu C. Hinoveanu, James Hopker, Bram Janssens, Hisanobu Kaji, Niklas Karlsson, Christos Katris, Beom-Jin Kim, Taeho Kim, Eiji Konaka, Michael A. Lapré, Anders Lunander, Daniel Memmert, Csaba Mihálykó, Éva Orbán-Mihálykó, Elizabeth M. Palazzolo, Anna Ibolya Pózna, Fumiya Shimizu, Benedek Soós, Tim B. Swartz, Dana Sylvan, Kazuya Takeda, Kazushi Tsutsui, Mert Türedioğlu, Rikuhei Umemoto, Fabian Wunderlich, Barbaros Yet, Tatsuya Yoshikawa

ISBN 978-963-503-941-8

Publisher: Corvinus University of Budapest

Budapest, 2023

Scientific Committee

Dries Goossens (Ghent University, Belgium)
Phil Scarf (Cardiff University, United Kingdom)
Marco Ferrante (University of Padova, Italy)
Dimitris Karlis (Athens University of Economics and Business, Greece)
Ruud Koning (University of Groningen, Netherlands)
Stephanie Kovalchik (Victoria University, Australia)
Ioannis Ntzoufras (Athens University of Economics and Business, Greece)
Alun Owen (Coventry University, United Kingdom)
James Reade (University of Reading, United Kingdom)
Frits Spieksma (TU Eindhoven, Netherlands)
Ray Stefani (California State University, Long Beach, United States)

Local Organising Committee

Chair

László Csató
Institute for Computer Science and Control (SZTAKI); Corvinus University of Budapest

Members

Sándor Bozóki
Institute for Computer Science and Control (SZTAKI); Corvinus University of Budapest

Dóra Gréta Petróczy
Institute of Economics, Centre for Economic and Regional Studies (KRTK KTI)
Corvinus University of Budapest

Zsombor Szádoczki
Institute for Computer Science and Control (SZTAKI); Corvinus University of Budapest

Preface

This is a collection of some papers to be presented at the 10th Mathsport International Conference, hosted by the Corvinus University of Budapest on June 26–28 June 2023. The 21 papers constitute short papers based around the presentation by one (or more) authors participating at the conference. Analogous to the conference, the papers presented here cover a wide range of sports, of methods, and of topics within sports. There are multiple papers on football, two papers on athletics, two on badminton, baseball, and cricket, as well as others on cycling, golf, tennis, and even on boulder climbing or team orienteering. The studies included in this volume show the increasing popularity of using sports data in empirical research. The papers are presented in no particular order.

Contents

Visual analysis of control area in badminton doubles using drone video dataset

Ning Ding[1,*], Kazuya Takeda[1], Yingjiu Bei[2], Keisuke Fujii[1,3,4]

[1] Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan
[2] Anhui Normal University, Wuhu, Anhui, China
[2] RIKEN Center for Advanced Intelligence Project, Fukuoka, Fukuoka, Japan
[3] PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan
[*] ding.ning@g.sp.m.is.nagoya-u.ac.jp

Abstract

Visual performance analysis of players in a dynamic competition is vital for effective sports coaching. In racket sports, most previous studies focused on analyzing and assessing singles players in broadcast videos and the discrete representations (e.g., stroke) that ignore meaningful spatial distributions. In our work, we built the first annotated drone dataset in badminton doubles and proposed a framework to estimate the control area probability map to evaluate teamwork performance in badminton doubles. In the experiment, we verified our approach by comparing various baselines and discovered the correlations between the score and control area. We also validated the effectiveness of our method through use cases. Our approach will provide coaches with valuable insights into doubles teamwork for assessing, recruiting, and scouting players.

1 Introduction

For sports analytics, advances in data collection have led to the further development of data-driven analysis of individuals and teams. One of the main goals of the game analysis is to provide visual and understandable answers to specific questions posed by the coach.

Significant effort has been focused on building larger broadcast sports video datasets [1, 4]. However, broadcast videos do not show the entire pitch, only provide partial information about the game, and are mostly available to wealthy professional sports teams. To be specific, there are frequent perspective changes in broadcast video. Besides, broadcast videos also suffer from severe occlusion, especially in team sports. In soccer, drone cameras are used to capture the entire pitch in a single frame [10] with the advantage of greater pitch adaptability and flexibility.

In this paper, we introduce a men's doubles badminton dataset captured with a 4K-resolution drone camera. The drone films the court from the top view, capturing the entire court without being affected by the occlusion problem.

Visual analytics has been widely used in various sports, such as heatmaps. In recent sophisticated methods, previous studies in team sports (e.g., soccer) aimed at estimating action (e.g., pass) probabilities and other performance metrics derived from spatiotemporal data. For example, a fully convolutional neural network architecture called SoccerMap estimates full probability surfaces of potential passes derived from high-frequency spatiotemporal data [11]. In racket sports, however, the task of fine-grained spatial analysis of game situations has received limited attention in prior research.

In racket sports, most previous studies focused on analyzing and assessing singles players using a single camera (e.g., [2, 13]) and the discrete representations (e.g., stroke) that ignore meaningful spatial distributions. Our research aims to estimate the control area probability map and provide coaches with insights into a team's control area as opponents always tend to attack the blind spots to win points. Although we may estimate various parameters of players such as visual angle and range of motions by deliberated experiments, we adopt a purely data-driven approach because we can usually access only game data. Developing such fine-grained analysis would facilitate coaches' decision-making in specific competitions.

The contributions of this paper are as follows:

- We build and share a men's doubles badminton drone dataset, annotated with bounding boxes and shuttlecock location (Hit/Drop), and with tracking algorithms.
- We propose a framework to estimate the control area probability map in a data-driven manner. This approach exploits spatial relationships and presents game situations in a visual format.
- In experiments, we verified our approach by comparing various baselines, and showed the correlation between the estimated control area and the score.

2 Dataset

2.1 Video collection

All data was obtained from 2-vs-2 badminton games between college badminton club members. Our data collection was conducted after receiving the approval of Anhui Normal University's ethics committee, and all participants provided signed informed consent.

We used a DJI Air 2S drone (Da-Jiang Innovations Science and Technology Co., Ltd., China) to capture the entire badminton court from the top view. The resolution was 4K (3,840 × 2,160 pixels), the frame rate was 30 fps and the raw video data was contains 39 games, covering 14 pairs, 11 players, and 1347 rallies.

2.2 Data annotation and structure

This section details our approach to efficiently annotating bounding boxes and the shuttlecock location in the drone video. Manual annotation is time-consuming that takes several hours to annotate a rally. Therefore, we utilized well-established computer vision techniques to shorten the process.

Here we outline the structural information of the dataset. Each rally is accompanied by 2 annotation files in simple comma-separated value (CSV) format. For shuttlecock detection, each line of the CSV file contains 5 values: frame, visibility, x-coordinate, y-coordinate, and status. For labeling, the status of the shuttlecock in each frame was annotated with one of the five types: Frying, Hit, Fault, Drop, and Misjudge. For players, we also provide corresponding bounding box coordinates in CSV format.

For tracking players and a shuttlecock from the raw video data, we first use homography transformation to eliminate the offset problem of drones. We segmented each game into several rallies. For each rally, we estimated the XY-coordinate values for the locations of 4 players using tracking algorithms (YOLO5 [6] and DeepSORT [12] with OSNet [14]), a popular high-precision object detection and tracking system. We detected the shuttlecock location using TrackNet [5]— an object tracking network proven to exhibit a decent tracking capability in games with tiny high-speed balls such as shuttlecock. We adjusted the outliers with a simple labeling tool using MATLAB GUI. By combining detection results with manual adjustment, we were able to speed up the annotation process significantly.

3 Visual analysis of control area

We propose a framework to estimate the control area probability map. To this end, we construct a neural network for learning the relationship between the tracking/event data and the map in a data-driven manner. In this section, we describe the neural network architecture and the learning method.

Table 1: Comparison of L1 classification loss for each input feature by eliminating the different input features from the total input.

Error	−Player's velocity	−Bbox height and width	Full
Control (hit)	0.145	0.132	**0.092**
Non-control (drop)	0.305	0.293	**0.252**
All	0.167	0.154	**0.101**

3.1 Model architecture

We constructed U-Net networks to estimate the control area full probability map. U-Net [9] is a convolutional network architecture for fast and precise segmentation of images. As the input, the networks take (1) A Gaussian mixed probability map centered on the location of 2 players (same team) who receive the shuttlecock, (2) X-velocity and Y-velocity of 2 players (same team) who receive the shuttlecock, (3) Width and height of the bounding box of 2 players (same team) who receive the shuttlecock. We take the location where the player hit the shuttlecock or dropped the shuttlecock as the target and finally obtain the control area probability map.

3.2 Learning

Our proposed loss function L consists of a focal loss and a constraint on spatial continuity, denoted as follows: $L = L_f + \mu L_c$, where μ represents the weight for balancing the two constraints, $\mu = 0.03$. As an objective function of our model, we first use Focal Loss [8] to address the class imbalance problem in our dataset such that $FL(p_{\mathrm{t}}) = -\alpha (1-p_{\mathrm{t}})^{\gamma} \log(p_{\mathrm{t}})$ where p_t is the estimated probability of the model. We set $\alpha = 0.8$ and $\gamma = 3$. The focal loss L_f can then be written as $L_f = FL\left(y_{loc_k}, f(x_k;\theta)_{loc_k}\right)$, where y is the probability of ground truth class, x_k is the game state at time k, loc_k is the location at time k where player hit the shuttlecock or the shuttlecock dropped. y_{lock} is the target at time k. In control area probability estimation, it is preferable that the probability of each pixel is spatially continuous. Similar to Kim et al. [7], we use an additional constraint L_c to suppress complicated patterns so we consider the L_1-norm difference in the horizontal and vertical direction. The spatial continuous loss is as follows: $L_c = \sum_{i=1}^{W-1} \sum_{j=1}^{H-1} \left\| v_{i+1,j} - v_{i,j} \right\|_1 + \left\| v_{i,j+1} - v_{i,j} \right\|_1$.

The network is trained using the Adam optimizer with a learning rate of 10^{-6}, 30 epochs, and batch sizes of 16. For the learning, we augmented our dataset with a horizontal flip. We have 12,658 hit samples and 796 drop samples. we used a ratio of 0.8 hit samples and 0.5 drop samples, and the rest for testing.

4 Results

In this section, we first verified the control area estimation model by visualizing the estimated control area and evaluating the accuracy. Second, we examined the practical usefulness of our approach by investigating the relationship between the control area and the score.

4.1 Control area estimation

First, we visualized the control area when a team reacts to an incoming shuttlecock after the shuttlecock crosses the net. The changes in the control area can help to understand the reason for the loss of points. The results presented the changes in the control area probability map of the receivers (both sides) during a catch in a rally. In this rally, receivers can hit (receive) the shuttlecock in cases of the first three results, while the last result shows that the left-side receivers failed to catch the shuttlecock. Therefore, the rally ended with a point lost by the left-side receiver. We did not make any assumptions regarding the shape of the control area distribution, which was learned from the data. We can also observe that the model may learn the player's speed for estimating the occupied spaces of the control area.

4.2 Verification of our method

Next, we evaluated our models using hit/drop samples and examined the impact of player velocity and the computed bounding box (Bbox) height and width, which approximates the player pose, on estimation performance. We computed the $L1$ classification loss between the ground truth and estimated positions. As presented in 1, For control (hit) and non-control (drop) samples, a model trained with all the input components (players' velocity, Bbox height and width) achieved the best performance, indicating that both the players' velocity and Bbox height and width information contributed to estimating an accurate control area probability map. For the model verification, our model achieved 0.101 for the test samples with hit and dropped shuttlecocks, where the losses of the hit and drop samples were 0.092 and 0.252, respectively.

4.3 Relationship with the score

4.3.1 Control area in the full field

First, we investigated the relationship between the score and the size of the control area on the full field. Here, 'full field' refers to one side of the output map that includes the area where a pair of players were located. We found no significant correlation between the score and the size of the control area ($p > 0.05$). We hypothesized that the size of the control area may be more related to the individual players' velocities than the team's performance (defense capabilities). Consequently, in the next section, we focused our analysis on the size of the control area in the primary area instead of the full field.

4.3.2 Control area of the primary field

Next, we analyzed the relationship between the score and the control area in the primary field and the proportion of the control area in order to evaluate the team's coverage performance (defense capabilities). Here, we defined the primary field as the field where the shuttlecock is located, which has a size of one-quarter of the output map size (56 × 28). Our results show a moderate positive monotonic correlation between the score and the control area in the primary field, with $\rho = 0.475$ ($p < 0.05$). We also found a similar correlation between the score and the proportion of the control area, with $\rho = 0.465$ ($p < 0.05$). These findings suggest that teams with better overall performance tend to have a larger control area in the primary field where the shuttlecock is located.

4.3.3 Length/width of the control area

In the game of doubles, "doubles rotation" is an important skill. The general idea is to cover the court and prevent any gaps during the match. For this reason, the coverage of the field can be used as an indicator of player performance.

We examined the control area of a team at the moment their opponents hit the shuttlecock. At that time, both players in the team should prepare and position themselves for the next stroke. The results showed that there was no correlation between the score and the length of the control area ($p > 0.05$). However, there was a weak positive monotonic correlation between the score and the width of the control area ($\rho = 0.226$, $p < 0.05$), suggesting that the team with better performance covers a wider area of the field when preparing for the next stroke.

5 Conclusion

In this study, we developed a framework to estimate the control area probability map from an input badminton drone video. We verified our approach by comparing various baselines and discovered valuable insights into the relationship between the score and control area. We shared the first annotated badminton drone video dataset and a tracking algorithm. Our approach can visually and quantitatively evaluate players' movements, which will be useful for coaching, recruitment, and scouting players. We expect this visual tool can be extended to other racket sports. In our future work, we plan to consider more dynamic indicators to reflect the skill including pose information [3] and extend our framework to a variety of other racket sports such as table tennis and tennis. We believe such visual representation of complex information can provide coaches with a deep perception of the game situation, thus providing a competitive advantage to an individual or a team.

Acknowledgments

This work was financially supported by JST SPRING, Grant Number JPMJSP2125, JSPS Grant Number 20H04075, JST PRESTO Grant Number JPMJPR20CA, and Scientific Research Project of Higher Education Institutions of Anhui Province Grant Number 2022AH052181. The author (ND) would like to take this opportunity to thank the "Interdisciplinary Frontier Next-Generation Researcher Program of the Tokai Higher Education and Research System." The authors would also like to thank Yundong Yu for his valuable comments on this work.

References

[1] A. Deliege, A. Cioppa, S. Giancola, M. J. Seikavandi, J. V. Dueholm, K. Nasrollahi, B. Ghanem, T. B. Moeslund, and M. Van Droogenbroeck. Soccernet-v2: A dataset and benchmarks for holistic understanding of broadcast soccer videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4508–4519, 2021.

[2] N. Ding, K. Takeda, and K. Fujii. Deep reinforcement learning in a racket sport for player evaluation with technical and tactical contexts. *IEEE Access*, 10:54764–54772, 2022.

[3] N. Ding, K. Takeda, W. Jin, Y. Bei, and K. Fujii. Estimation of control area in badminton doubles with pose information from top and back view drone videos. *arXiv preprint arXiv:2305.04247*, 2023.

[4] S. Giancola and B. Ghanem. Temporally-aware feature pooling for action spotting in soccer broadcasts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2021.

[5] Y.-C. Huang, I.-N. Liao, C.-H. Chen, T.-U. İk, and W.-C. Peng. Tracknet: A deep learning network for tracking high-speed and tiny objects in sports applications. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8. IEEE, 2019.

[6] G. Jocher, A. Stoken, J. Borovec, S. Christopher, and L. C. Laughing. ultralytics/yolov5: v4. 0-nn. silu () activations, weights & biases logging, pytorch hub integration. *Zenodo*, 2021.

[7] W. Kim, A. Kanezaki, and M. Tanaka. Unsupervised learning of image segmentation based on differentiable feature clustering. *IEEE Transactions on Image Processing*, 29:8055–8068, 2020.

[8] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[9] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th*

International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pages 234–241. Springer, 2015.

[10] A. Scott, I. Uchida, M. Onishi, Y. Kameda, K. Fukui, and K. Fujii. Soccertrack: A dataset and tracking algorithm for soccer with fish-eye and drone videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3569–3579, 2022.

[11] T. Taki and J.-i. Hasegawa. Visualization of dominant region in team games and its application to teamwork analysis. In *Proceedings computer graphics international 2000*, pages 227–235. IEEE, 2000.

[12] N. Wojke, A. Bewley, and D. Paulus. Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)*, pages 3645–3649. IEEE, 2017.

[13] T. Yoshikawa, N. Ding, K. Takeda, and K. Fujii. Prediction of shot type and hit location based on pose information using badminton match videos. *MathSport International*, 2023.

[14] K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang. Omni-scale feature learning for person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3702–3712, 2019.

Prediction of shot type and hit location based on pose information using badminton match videos

Tatsuya Yoshikawa[1†], Ning Ding[1†], Kazuya Takeda[1], Keisuke Fujii[1,2,3,*]

[1] Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan
[2] RIKEN Center for Advanced Intelligence Project, Fukuoka, Fukuoka, Japan
[3] PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan
[†] equal contribution
[*] fujii@i.nagoya-u.ac.jp

Abstract

With the development of video processing technology, it is expected to be used for various motion analysis in racket sports. The previous badminton model predicted stroke types and hitting positions based on match situations and player position information. However, prediction without motion information is difficult to apply to actual play. In this study, we propose a multi-task deep learning model that predicts the stroke type and hitting location based on the estimated pose information from badminton singles match videos. Experimental results show that the proposed method predicted the stroke type and hitting position more accurately than the existing method without pose. Furthermore, by introducing a mechanism called a gradient reversal layer to decrease the bias between the front and the back of the image, the proposed method predicted more accurately the stroke type. Our approach is expected to be used in games and practices by understanding easily predictable movements.

1 Introduction

Advancements in video analysis technology have paved the way for its application in various motion analyses for racket sports. Practical el players aim to strike the ball in advantageous situations, anticipating their opponent's next stroke type and hit location according to the game situations. In other words, favorable game progression is expected by predicting the opponent's subsequent shot based on their movements. For instance, a badminton motion analysis study [8] focuses on racket angles and velocities. However, the duration of a ball strike is incredibly brief, making it impossible to visually track the racket's information at the moment of impact. As a result, it is more desirable to analyze human joint movements (e.g., [13, 3, 4]) since predictions are often based on observing body movements rather than the racket.

In recent years, several research studies have employed machine learning to predict stroke types and hitting locations. ShuttleNet [12] predicted the stroke type and the location of the hit ball from the game situation such as rallies and scores, and player information such as the locations of both players. However, it is difficult to make use of predictions without movement information in a practical sense. Meanwhile, in another work [2], the stroke type was classified based on the pose information, which indicates that the pose information can be used for classifying the stroke type.

In this study, we propose a multi-task deep learning model to predict the stroke type and the hitting locations using machine learning based on pose information from badminton singles match videos. In addition, we introduce a mechanism called Gradient Reversal Layer (GRL) [7] to reduce the bias of pose information in the front and back of the image, which can be a problem for monocular images. The experimental results demonstrated the superiority of the proposed method compared to existing methods that lack posture information and GRL. This method is expected to be useful for games and practice, as it enables the prediction of movements that are easy for humans to anticipate.

2 Methods

2.1 Dataset

In this research, four badminton players with 7.5 ± 1.8 years of experience (mean $\pm$ standard deviation) participated in the measurement of badminton games. Since the angle of view often seen in professional tournaments and filming from upper spectator seats are common, we positioned the camera facing the front (back) of the athletes from a high vantage point.

We filmed 14 five-minute singles games, resulting in a total score of 362 points and 1664 rallies. The games were arranged in a round-robin format so that each of the four players played against each other, with the direction of the players facing the camera alternating between games. We used a digital camera (RX02, Sony, Japan) to capture videos at a frame rate of 60 fps (frames per second). Before participating in the experiment, all participants received a full explanation of the study and provided written informed consent. All experimental procedures were performed with prior approval from the Graduate School of Informatics, Nagoya University, for experiments involving human subjects.

2.2 Joint Position Coordinate Estimation

In this study, we used the model HigherHRNet [1][9][11] to estimate joint position coordinates from badminton match videos. Although the number of estimation points can be selected, in this study, due to the long distance between the camera and the player and the short computation time, we simply used both shoulders, both elbows, both wrists, and both hips, both knees, and both ankles (total 12 points). During the estimation process, the estimation points of both players, especially the player who is farther from the camera, were occasionally missing. To address this issue, we introduced two-dimensional linear interpolation to closely approximate the missing points.

In addition, both x and y coordinates were normalized as data reduction for prediction. A pseudo bounding box was formed by taking the maximum and minimum values of the joint points for each frame. Based on those values, the x-coordinate was normalized to a value between -1 and 1, and the y-coordinate was also normalized while preserving the aspect ratio to maintain the human form.

2.3 Stroke Type and Strike Position Labeling

As training data, we labeled the frames containing shots in each video. We manually labeled 10 stroke types, which include serve, smash, drop, clear, lift, drive, push, net shot, net kill, and block. Regarding the hitting position, the positions of both feet of the player were obtained from HigherHRNet, and then the midpoint was calculated. Two-dimensional Direct Linear Transformation (DLT) was performed with the lower left of the court as (0, 0) and the upper right as (610, 1340) [cm]. For predictions, the center of the court was taken as the origin and normalized to [-1,1].

3 ShuttlePoseNet

In this study, we constructed a prediction model based on ShuttleNet's LSTM model [12] by adding the pose embedding module. The input information consists of the joint position coordinates, hitting position coordinates, and stroke type at the time of hitting the shuttlecock. LSTM (Long Short Term Memory) is an improved version of RNN (Recurrent Neural Network) that takes into account time series and solves the gradient vanishing problem of RNN.

The encoder is based on LSTM, and the decoder is based on LSTM and GRL (Gradient Reversal Layer) [7]. The coordinates of the strike positions of the players are mainly concentrated in two places across the net, which led us to believe that the court where the shuttlecock was hit could bias the prediction. In

GRL [7] (used human movements in [6, 5], the forward propagation of the input remains unchanged, while the backpropagation reverses the gradient by multiplying it with a negative scalar. To reduce the bias, we introduced a layer of GRL just before the layer that predicts which side of the court the player hit the shuttlecock.

Finally, we made predictions and obtained three outputs: (1) The sampled hitting position coordinates from the normal distribution for the stroke after the $(n+1)$-th stroke, (2) the stroke type, and (3) whether the ball was hit by the player on the back or front side of the court.

The total loss function $\mathscr{L}$ was computed as follows:

$$\mathscr{L} = \mathscr{L}_{type} + \mathscr{L}_{player} + \mathscr{L}_{area}. \tag{1}$$

We trained our model using a multi-task learning approach, where we jointly optimized three losses: $\mathscr{L}_{type}$, which measures the prediction error of stroke type; $\mathscr{L}_{player}$, which measures the prediction error of the hitting player; and $\mathscr{L}_{area}$, which measures the prediction error of the hitting position area.

For learning, we use 4 observed strokes as input for each rally, and training is carried out with 80% of the dataset while evaluation is conducted using the remaining 20%. Rallies with fewer than 5 strokes were excluded from the analysis to avoid predicting them in the first place.

4 Results

In this section, we conduct evaluation experiments on the proposed method, ShuttlePoseNet. We divided all the data into 5 parts and performed 5-fold cross-validation. From the results, we calculated the mean and standard deviation. We compared our full model (ShuttlePoseNet) with three baselines: the LSTM-only model, the LSTM-only model without pose information, and the LSTM+GRL model without pose information. LSTM-only model is the LSTM version of ShuttleNet [12] as the comparable baseline. In addition, we changed the input shot number n in the encoder/decoder and compared their performances. As evaluation metrics, we used accuracy for stroke type prediction, error distance between the predicted and true values for hitting position prediction, and hit position accuracy based on the court division. The court was divided into 20 sections based on the lines of the badminton court. The reason for dividing the court into meshes is that players do not always hit the ball in the same position when returning it to the same spot. Considering that the hitting position may deviate to some extent, it is reasonable to divide the court into meshes.

4.1 Prediction performances

We performed a quantitative evaluation of predictive performance. Table 1 shows the comparison results of each model. For reference, the court size is 610 $\times$ 1340 cm. Prediction with pose information was more accurate for both stroke type and hit position, whereas vice versa in the hit position error. This suggests that while the small distance errors increased, global errors (as areas) were reduced. Furthermore, LSTM+GRL was more accurate than the LSTM-only model. These suggest that our approach improved the prediction performances compared to the baselines.

4.2 Time length of input data

In this study, the ShuttlePoseNet model takes input data from the serve shot up to the τ-th shot and uses it as input for the encoder, while predicting the subsequent shots using the decoder. We compared the results by varying the number of shots used as input for the encoder and decoder, and the results are presented in Table 2. In all cases, the predictions were made for 4 shots after input. The most accurate predictions were obtained with $\tau = 3,4$. Since the service shot is at $\tau = 1$ and the subsequent return shot is at $\tau = 2$, and

	Stroke type accuracy (%)	hit position error (cm)	hit position accuracy (%)
LSTM (w/o pose)	16.08 ± 0.008	86.98 ± 2.461	86.43 ± 0.009
LSTM	17.09 ± 0.708	89.43 ± 2.062	86.44 ± 0.463
LSTM+GRL (w/o pose)	15.58 ± 1.325	87.07 ± 2.437	86.44 ± 1.161
ShuttlePoseNet	17.53 ± 1.224	96.24 ± 1.266	86.65 ± 0.956

Table 1: Comparison of preidciton performances for various models.

	Stroke type accuracy (%)	hit position error (cm)	hit position accuracy (%)
$\tau = 2$	17.43 ± 1.019	95.65 ± 0.901	84.40 ± 1.150
$\tau = 3$	18.03 ± 1.174	92.71 ± 0.531	85.38 ± 0.803
$\tau = 4$	17.53 ± 1.224	96.24 ± 1.266	86.65 ± 0.956
$\tau = 5$	17.14 ± 1.387	103.60 ± 1.695	84.30 ± 1.129

Table 2: Validation of the models when changing the time length τ.

their positions are almost the same, it may be possible to predict the approximate flow of the game in terms of stroke type and hitting position for the next 2-3 shots. However, with more data, it is possible that the outcome of the point has already been determined, and there may be insufficient data for accurate prediction. This suggests that there may not be enough data for sufficient learning beyond a certain point.

5 Conclusion

In this study, we proposed a multi-task deep learning model that estimates the joint position coordinates from badminton singles match videos and predicts the stroke type and the hit position by machine learning based on the information. In the experiment, the proposed method with posture information was able to predict the stroke type and hitting position more accurately than the existing method without posture information. Furthermore, by introducing a mechanism called a gradient reversal layer to remove the bias in the foreground and the background of the image, we were able to predict the stroke type more accurately. This method is expected to be useful in games and practices because it enables us to predict movements that are easy for humans to anticipate.

In this study, we used only 1664 strokes from videos, which we record by ourselves as the dataset. In a deep learning model, the number of data points is an important factor in determining accuracy. Therefore, an increase in the amount of data is expected to lead to improved accuracy. For example, since the angle of view in the videos we obtained was similar to that of publicly available videos on YouTube and TV, it is possible to extend the data by annotating and using public data. In addition, the input of pose information is limited to one frame at the moment of hitting the shuttlecock, but since a series of joints is important for the flow of movement, using pose information from all frames between hits is likely to improve accuracy. Furthermore, to improve accuracy in joint position coordinate estimation, it would be beneficial to perform fine-tuning of the pretrained pose estimation model (e.g., [10]).

Acknowledgments

This work was financially supported by JSPS Grant Number 20H04075, JST PRESTO Grant Number JPMJPR20CA, and JST SPRING, Grant Number JPMJSP2125. The author (ND) would like to take this opportunity to thank the "Interdisciplinary Frontier Next-Generation Researcher Program of the Tokai Higher Education and Research System."

References

[1] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation. In *CVPR*, 2020.

[2] W.-T. Chu and S. Situmeang. Badminton video analysis based on spatiotemporal and stroke features. In *Proceedings of the 2017 ACM on international conference on multimedia retrieval*, pages 448–451, 2017.

[3] N. Ding, K. Takeda, and K. Fujii. Deep reinforcement learning in a racket sport for player evaluation with technical and tactical contexts. *IEEE Access*, 2022.

[4] N. Ding, K. Takeda, W. Jin, Y. Bei, and K. Fujii. Estimation of control area in badminton doubles with pose information from top and back view drone videos. *arXiv preprint arXiv:2305.04247*, 2023.

[5] K. Fujii, K. Takeuchi, A. Kuribayashi, N. Takeishi, Y. Kawahara, and K. Takeda. Estimating counterfactual treatment outcomes over time in complex multi-agent scenarios. *arXiv preprint arXiv:2206.01900*, 2022.

[6] K. Fujii, K. Takeuchi, A. Kuribayashi, N. Takeishi, Y. Kawahara, and K. Takeda. Estimating counterfactual treatment outcomes over time in multi-vehicle simulation. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, pages 1–4, 2022.

[7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

[8] M. Kwan, M. S. Andersen, C.-L. Cheng, W.-T. Tang, and J. Rasmussen. Investigation of high-speed badminton racket kinematics by motion capture. *Sports Engineering*, 13(2):57–63, 2011.

[9] K. Sun, B. Xiao, D. Liu, and J. Wang. Deep high-resolution representation learning for human pose estimation. In *CVPR*, 2019.

[10] T. Suzuki, K. Takeda, and K. Fujii. Automatic fault detection in race walking from a smartphone camera via fine-tuning pose estimation. In *2022 IEEE 11th Global Conference on Consumer Electronics (GCCE)*, pages 631–632. IEEE, 2022.

[11] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao. Deep high-resolution representation learning for visual recognition. *TPAMI*, 2019.

[12] W.-Y. Wang, H.-H. Shuai, K.-S. Chang, and W.-C. Peng. Shuttlenet: Position-aware fusion of rally progress and player styles for stroke forecasting in badminton. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4219–4227, 2022.

[13] T. Yoshikawa, K. Tsutsui, K. Takeda, and K. Fujii. Extraction of swing motion contributing to prediction of shuttle drop position in badminton. *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI'21) Workshop on AI for Sports Analytics (AISA)*, 8 2021.

Bike2Vec: Vector embedding representations of road cycling riders and races

Ethan Baron[a], Bram Janssens[b,c,d], Matthias Bogaert[b,c]

[a] University of Toronto / Zelus Analytics, eth.baron@mail.utoronto.ca
[b] Department of Marketing, Innovation and Organisation, Ghent University
[c] FlandersMake@UGentâcorelab CVAMO
[d] Research Foundation Flanders

Abstract

Vector embeddings have been successfully applied in several domains to obtain effective representations of non-numeric data which can then be used in various downstream tasks. We present a novel application of vector embeddings in professional road cycling by demonstrating a method to learn representations for riders and races based on historical results. We use unsupervised learning techniques to validate that the resultant embeddings capture interesting features of riders and races. These embeddings could be used for downstream prediction tasks such as early talent identification and race outcome prediction.

1 Introduction

Professional road cycling offers several interesting challenges in an analytics setting due to its unique properties. Notably, races have a variety of different formats (e.g. one-day races, stage races) and profiles (e.g. flat, hilly, or mountainous), each suiting riders of different characteristics. Although several past works have demonstrated the potential of using machine learning models to predict road cycling race results, these models rely on significant feature engineering efforts and are tailored to predicting specific outcomes, such as rider performance in a specific race.

We present a framework forming the foundation for a generalized prediction algorithm that does not depend on labour-intensive feature engineering efforts. Specifically, we introduce a method to train vector embeddings for riders and races based on historical race results.

In representation learning, vector embeddings are used to capture the key qualities of entities such as words, images, or songs. If trained effectively, these vector embeddings can then be used for a variety of downstream tasks. For example, word embeddings trained using a large corpus of text can be used for emotion recognition or sentence completion.

Likewise, we show that our cycling embeddings capture the key characteristics of riders and races. The embeddings can be used in downstream prediction tasks and eliminate the need for a manual feature engineering process.

2 Literature Review

2.1 Machine Learning in Road Cycling

There are several prior works which apply machine learning to road cycling.

Multiple works focus on predicting the ProCyclingStats ("PCS") points, a system developed by the website `procyclingstats.com` to assign scores to riders based on the results achieved in certain races. For example, Janssens and Bogaert (2021) construct a random forest regression to predict the total PCS points scored by under-23 prospects in their first two years as professional athletes. They engineer a large set of features, including the riders' performances in particular under-23 races, and compare various methods to impute non-participated race results. This imputation method is used to detect the most promising young athletes (Janssens, Bogaert, and Maton 2022). Similarly, Van Bulck, Vande Weghe, and Goossens (2021) compare linear regression and random forest regression to predict the points scored by under-23 riders in their first three years as professionals. They also hand-craft a number of features summarizing the riders' performance at the youth level.

Other works focus on predicting the outcomes of particular races. De Spiegeleer (2019) develops machine learning models to predict various outcomes of stages from the Tour de France, Giro d'Italia, and Vuelta a Espana, including their average speed, the difference between the average speed of the winner and that of a particular rider, and the head-to-head performance between two riders. The predictions are based on an extensive set of engineered features related to the terrain, weather, rider characteristics, and historical results. Mortirolo (2019) uses Bayesian Additive Regression Trees to simulate races and obtain predictions for the probabilities of specific race outcomes. The simulation uses over one hundred features, including ratings for riders' various abilities, indicators of riders' recent form, historical results from the past three years, and team-level indicators. Kholkine, De Schepper, et al. (2020) apply an XGBoost model to predict the outcomes of the Tour of Flanders using riders' performances in relevant build-up races. They also engineer several advanced features related to past results in similar races, historical weather data and overall team performance. Kholkine, Servotte, et al. (2021) also employ an XGBoost model within a learn-to-rank framework to predict the top ten riders in several one-day races using a suite of engineered features based on historical results and the riders' ages. Finally Demunter (2021) compares linear regression, random forest, XGBoost, and neural networks to predict the rankings of riders in a given race. Again, various features related to the rider's recent and historical results are developed and used as inputs for these models.

To the best of our knowledge, we present the first framework for a generalized prediction algorithm for road cycling which does not rely on a hand-crafted set of features for the particular outcome of interest.

2.2 Representation Learning

Representation learning is the field of machine learning concerned with automatically learning meaningful and compact representations of data without requiring features for the data. These representations aim to capture the underlying structure and patterns in the data, enabling more effective performance on a variety of downstream applications. Some primary applications of representation learning include natural language processing, computer vision, and recommendation systems.

In natural language processing, representation learning has been utilized to learn word embeddings that capture semantic relationships between words. For example, the word2vec algorithm uses a skip-gram approach to fit vector embeddings for words. These embeddings yield successful results on downstream tasks such as semantic and syntactic word relationship testing and sentence completion (Mikolov et al. 2013). Another common approach known as GloVe uses co-occurence statistics of words to obtain the vector embeddings. The resulting representations performed strongly on a variety of tasks including word analogies, word similarities, and named entity recognition (Pennington, Socher, and Manning 2014). Finally, pre-trained vector embeddings using bidirectional encoder representations from transformers (BERT) can then be used as inputs into various downstream tasks and have yielded state-of-the-art performance on question answering and language inference (Devlin et al. 2019).

Similarly, representation learning has been successfully applied in the field of computer vision. One notable technique is the use of convolutional neural networks. These convolutional neural networks achieved pioneering performance on image classification tasks, such as handwritten digit recognition (Cireşan et al. 2011) and high-resolution image classification (Krizhevsky, Sutskever, and Hinton 2012). By pre-training large convolutional neural networks on large amounts of data, researchers have then achieved state-of-the-art performance on novel computer vision tasks, including image recognition (Simonyan and Zisserman 2015), object detection (Girshick et al. 2014), scene recognition and domain adaptation (Donahue et al. 2014).

Representation learning advancements in natural language processing and computer vision have exploited observed relationships between words and local patterns in images. In the case of road cycling, we seek to extract representations for both races and riders and exploit historical interactions between these two types of entities. Most relevant to this context are past works on recommender systems surrounding collaborative filtering, which use historical interactions between users and items to recommend new items to a user. One common approach for such problems is to transform both items and users to the same vector space by assigning vector embeddings of the same dimension to both categories. Then, dot products between the vector embeddings of users and items can be used to infer their interaction (Koren 2008).

Although we are not aware of such a representation learning approach being applied in road cycling, similar approaches have recently been tested in other sports, including soccer (Cintia and Pappalardo 2021; Magdaci 2021; Yılmaz and Öğüdücü 2022; Li et al. 2022), basketball (Papalexakis and Pelechrinis 2018), and cricket (Alaka, Sreekumar, and Shalu 2021).

3 Methods

We collect historical race results from the 2016-2022 UCI World Tour seasons from `procyclingstats.com`. Specifically, we consider the results of one-day races, and individual stages of stage races (i.e. multi-day races), except for team time trials. We do not consider overall classifications of stage races. Overall, our dataset includes results from 1069 race editions, 118 of which are one-day races.

For each result in our data, we define the *normalized PCS score* as the number of PCS points scored by the rider in the race, divided by the points earned by the winner of that race. For example, if the winner and runner-up of a race earn 500 and 300 PCS points respectively, they are assigned a normalized PCS score of 1 and 0.6 respectively.

We learn vector embeddings of dimension D for individual riders and races by directly optimizing these embeddings' ability to predict historical results. Specifically, we represent a rider's aptitude at a given race by the dot-product between that rider's embedding and that race's embedding. We then pass this dot-product through a sigmoid activation function to predict the normalized PCS score for that rider at that race. The vector embeddings are trained by minimizing the binary cross-entropy loss between these predictions and the actual normalized PCS scores, according to equation 1.

$$L(R,S,y) = \frac{1}{N}\sum_{i=1}^{N} y_i \log(\sigma(R_{r(i)} \cdot S_{s(i)})) + (1-y_i)\log(1-\sigma(R_{r(i)} \cdot S_{s(i)})) \tag{1}$$

Here, y_i records the normalized PCS points scored by rider $r_{(i)}$ in race $s_{(i)}$, R is the matrix of rider vector embeddings, S is the matrix of race vector embeddings, and N is the number of results in our data. σ refers to the sigmoid function such that $\sigma(x) = (1+e^{-x})^{-1}$.

We train a vector embedding for each rider who has scored at least 25 (unnormalized) PCS points in our dataset. We also train a vector embedding for each race edition, except that for one-day races, we use the same embedding across all seasons. We do this since one-day races tend to suit similar riders across years, whereas stages can have very different characteristics across seasons. In total, we train unique embeddings for 973 races and 958 riders.

The results shown below are based on embeddings of dimension $D = 5$ trained using an Adam optimizer with a learning rate of 0.001 for 100 epochs. Reproducible code to implement our methods is available at `https://github.com/baronet2/Bike2Vec`.

4 Results

In this section, we analyze our learned embeddings to demonstrate that they capture valuable information about the characteristics of riders and races.

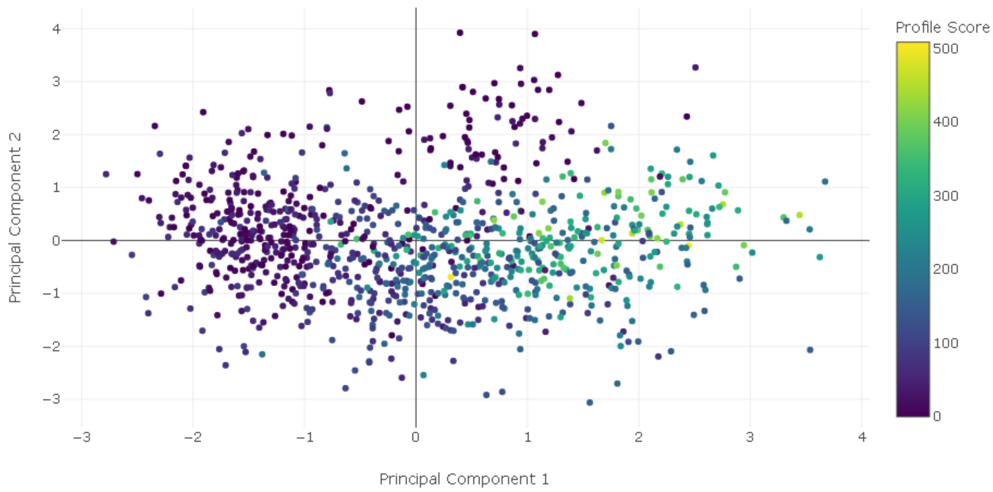


Figure 1: Visualization of race embeddings coloured by race profile score.

In Figure 1, we plot the embeddings for each race in our dataset, coloured according to the race profile score, a measurement of the amount of climbing in the race developed by PCS. We performed a principal component analysis to reduce the dimensionality of the embeddings to two dimensions for visualization purposes. Clearly, the primary principal component is capturing significant information about the terrain of a race, with more mountainous races appearing on the right and flat races appearing on the left.

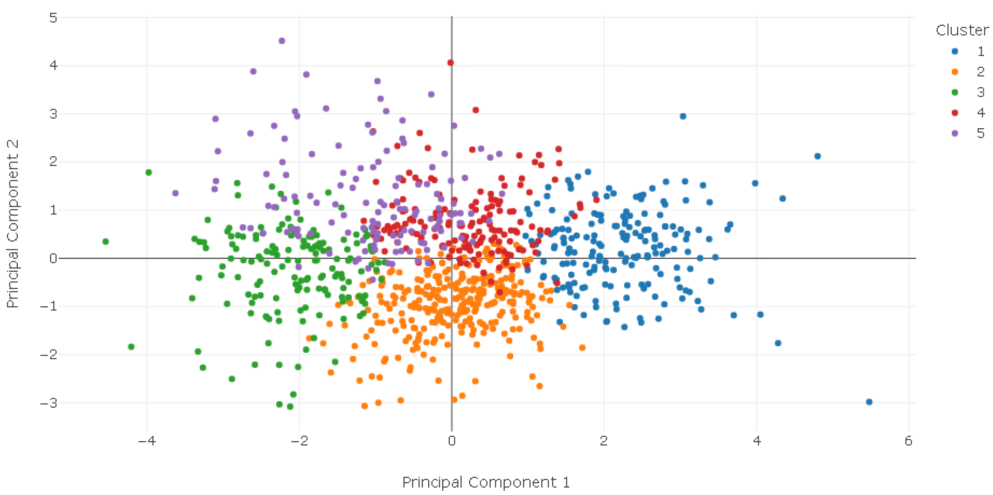


Figure 2: Visualization of rider embeddings coloured by cluster.

Similarly, in Figure 2, we plot the principal components of the rider embeddings. Unlike races, riders are not labelled by PCS as belonging to a certain category. Therefore, to add interpretability, we perform k-means clustering on the rider embeddings and colour the riders by their assigned cluster.

We show a few examples of riders from each cluster in Table 1. There are clear similarities among the riders in each cluster, indicating that our embeddings are capturing the unique characteristics of each rider.

For example, cycling fans would identify that cluster 1 is composed of sprinters and cluster 3 of climbers.

Cluster	Examples of Riders
1	SAGAN Peter, KRISTOFF Alexander, VIVIANI Elia, EWAN Caleb, BENNETT Sam
2	VAN AVERMAET Greg, COLBRELLI Sonny, NAESEN Oliver, MOHORIČ Matej
3	ALAPHILIPPE Julian, VALVERDE Alejandro, ROGLIČ Primož, POGAČAR Tadej
4	VAN AERT Wout, MATTHEWS Michael, STUYVEN Jasper, KWIATKOWSKI Michałl
5	VAN DER POEL Mathieu, GILBERT Philippe, LAMPAERT Yves, Š TYBAR Zdeněk

Table 1: Examples of riders from each cluster.

Rider 1	Rider 2
POGAČAR Tadej	ROGLIČ Primož
SAGAN Peter	COLBRELLI Sonny
ALAPHILIPPE Julian	HIRSCHI Marc
YATES Simon	BARDET Romain
EVENEPOEL Remco	ALMEIDA João
QUINTANA Nairo	ZAKARIN Ilnur
VIVIANI Elia	GREIPEL André
DENNIS Rohan	CAVAGNA Rémi

Table 2: Examples of most similar rider embeddings.

Furthermore, in Table 2, we show some examples of rider similarities. That is, for each rider on the left-hand-side, we show the name of the other rider with the most similar embedding, according to Euclidean distance. Cycling fans would confirm that these rider pairings strongly reflect these riders' characteristics. For example, Tadej Pogačar and Primož Roglič are both world-class climbers and time-trialists, Peter Sagan and Sonny Colbrelli are versatile sprinters who also perform well in cobbled or hilly classics, and Julian Alaphilippe and Marc Hirschi are both specialists at climbing short but steep hills.

Overall, the vector embeddings seem to accurately capture the distinguishing characteristics of both riders and races.

5 Conclusion

We present a novel vector embedding approach to represent road cycling riders and races, and implement the approach on seven seasons of data from professional men's road cycling races. We validate the resulting embeddings by showing that they contain useful information about the characteristics of races and riders. These embeddings can form the basis for a variety of downstream prediction tasks, removing the need for extensive manual feature engineering.

Although we have demonstrated that our proposed vector embeddings contain valid information about the riders and races, we have yet to test the inclusion of these embeddings within a downstream prediction task. We leave this as a promising area for future work. Further, augmenting our race embeddings using features about the route, such as the elevation profile, could offer improved race embeddings and enable predictions on new races. Additionally, our current framework assigns a constant embeddings over the span of riders' careers. Future research could seek to incorporate a time-varying element to capture changes in rider skills due to age, physiology, injury, or other effects. Lastly, an interesting avenue for further exploration is extending our framework to women's cycling and comparing the results.

References

Alaka, Souridas, Rishikesh Sreekumar, and Hrithwik Shalu (Aug. 2021). *Efficient Feature Representations for Cricket Data Analysis using Deep Learning based Multi-Modal Fusion Model*. eprint: 2108.07139. URL: https://arxiv.org/pdf/2108.07139.pdf.

Cintia, Paolo and Luca Pappalardo (June 2021). *Coach2vec: Autoencoding the Playing Style of Soccer Coaches*. Tech. rep. ISTI Research Report. eprint: 2106.15444. URL: https://arxiv.org/ftp/arxiv/papers/2106/2106.15444.pdf.

Cireşan, Dan Claudiu et al. (July 2011). "Flexible, High Performance Convolutional Neural Networks for Image Classification". In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*. Vol. 2. Ijcai'11. Barcelona: AAAI Press, pp. 1237–1242. ISBN: 9781577355144. URL: https://people.idsia.ch/~juergen/ijcai2011.pdf.

De Spiegeleer, Emiel (June 2019). "Predicting Cycling Results using Machine Learning". MA thesis. Ghent: Ghent University. URL: https://libstore.ugent.be/fulltxt/RUG01/002/785/834/RUG01-002785834%5C%5F2019%5C%5F0001%5C%5FAC.pdf.

Demunter, Jarne (June 2021). "Predicting Ranking Multientrant Races: Road Cycling". MA thesis. Ghent: Ghent University. URL: https://libstore.ugent.be/fulltxt/RUG01/003/010/353/RUG01-003010353%5C%5F2021%5C%5F0001%5C%5FAC.pdf.

Devlin, Jacob et al. (Oct. 2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. eprint: 1810.04805. URL: https://arxiv.org/pdf/1810.04805.pdf.

Donahue, Jeff et al. (June 2014). "DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition". In: *Proceedings of the 31st International Conference on Machine Learning*. Ed. by Eric P. Xing and Tony Jebara. Vol. 32. Proceedings of Machine Learning Research 1. Bejing, China: Pmlr, pp. 647–655. URL: https://proceedings.mlr.press/v32/donahue14.html.

Girshick, Ross et al. (June 2014). "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos: IEEE Computer Society, pp. 580–587. DOI: 10.1109/cvpr.2014.81.

Janssens, Bram and Matthias Bogaert (Sept. 2021). "Imputation of Non-participated Race Results". In: *ECML / PKDD 2021 Workshop, 8th Workshop on Machine Learning and Data Mining for Sports Analytics, Proceedings*. Bilbao, p. 12. URL: https://dtai.cs.kuleuven.be/events/MLSA21/papers/MLSA21%5C%5Fpaper%5C%5Fjanssens.pdf.

Janssens, Bram, Matthias Bogaert, and Mathijs Maton (Jan. 2022). "Predicting the next Pogačar: a data analytical approach to detect young professional cycling talents". In: *Annals of Operations Research*, pp. 1–32. DOI: https://doi.org/10.1007/s10479-021-04476-4.

Kholkine, Leonid, Tom De Schepper, et al. (Sept. 2020). "A Machine Learning Approach for Road Cycling Race Performance Prediction". In: *Machine Learning and Data Mining for Sports Analytics*. Ed. by Ulf Brefeld et al. Communications in Computer and Information Science. Springer International Publishing, pp. 103–112. ISBN: 978-3-030-64912-8. DOI: 10.1007/978-3-030-64912-8_9.

Kholkine, Leonid, Thomas Servotte, et al. (Oct. 2021). "A Learn-to-Rank Approach for Predicting Road Cycling Race Outcomes". In: *Frontiers in Sports and Active Living* 3, p. 714107. ISSN: 2624-9367. DOI: 10.3389/fspor.2021.714107.

Koren, Yehuda (Aug. 2008). "Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model". In: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discov-*

ery and Data Mining. Kdd '08. Las Vegas: Association for Computing Machinery, pp. 426–434. ISBN: 9781605581934. DOI: 10.1145/1401890.1401944.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton (Dec. 2012). "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by F. Pereira et al. Vol. 25. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper%5C%5Ffiles/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Li, Yuesen et al. (July 2022). "Characterizing Player's Playing Styles Based on Player Vectors for Each Playing Position in the Chinese Football Super League". In: *Journal of Sports Sciences* 40.14, pp. 1629–1640. DOI: 10.1080/02640414.2022.2096771.

Magdaci, Ofir (June 2021). *Embedding the Language of Football Using NLP*. URL: https://towardsdatascience.com/embedding-the-language-of-football-using-nlp-e52dc153afa6.

Mikolov, Tomas et al. (Jan. 2013). *Efficient Estimation of Word Representations in Vector Space*. eprint: 1301.3781. URL: http://arxiv.org/abs/1301.3781.

Mortirolo (July 2019). *Cycling prediction method*. URL: https://mortirolo.netlify.app/post/cycling-prediction-method/ (visited on 05/06/2023).

Papalexakis, Evangelos and Konstantinos Pelechrinis (Oct. 2018). "THoops: A Multi-Aspect Analytical Framework for Spatio-Temporal Basketball Data". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. Cikm '18. Torino: Association for Computing Machinery, pp. 2223–2232. ISBN: 9781450360142. DOI: 10.1145/3269206.3272002.

Pennington, Jeffrey, Richard Socher, and Christopher Manning (Oct. 2014). "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha: Association for Computational Linguistics, pp. 1532–1543. DOI: 10.3115/v1/D14-1162.

Simonyan, Karen and Andrew Zisserman (Apr. 2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *3rd International Conference on Learning Representations*. San Diego: Computational and Biological Learning Society, pp. 1–14. eprint: 1409.1556. URL: https://arxiv.org/pdf/1409.1556.pdf.

Van Bulck, David, Arthur Vande Weghe, and Dries Goossens (Oct. 2021). "Result-based talent identification in road cycling: Discovering the next Eddy Merckx". In: *Annals of Operations Research*. DOI: 10.1007/s10479-021-04280-0.

Yılmaz, Öznur Ilayda and Şule Gündüz Öğüdücü (Apr. 2022). "Learning Football Player Features Using Graph Embeddings for Player Recommendation System". In: *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*. Sac '22. Association for Computing Machinery, pp. 577–584. ISBN: 9781450387132. DOI: 10.1145/3477314.3507257.

Applications of point pattern analysis in baseball

Hisanobu Kaji* and Dana Sylvan**

* 695 Park Avenue, New York, NY 10065, USA, hisanobu.kaji32@myhunter.cuny.edu
** 695 Park Avenue, New York, NY 10065, USA, dsylvan@hunter.cuny.edu

Abstract

Statistics is an integral part of baseball, all the more in the last years with the huge developments in technology and data science. A wealth of baseball data is freely available and contains spatial information connected to location coordinates for individual batted balls captured by high-speed cameras. In this project we focus on modeling and visualization of baseball data via point pattern analysis. We use kernel smoothing to produce spatial heat maps of batted balls for top Major League Baseball players. In particular, we compare and contrast patterns produced by right-handed, left-handed, and switch hitters, respectively. After rejecting the hypothesis of complete spatial randomness we employ kernel smoothing to model and visualize the space-varying intensity function of the resulting inhomogeneous Poisson process. The procedure includes a routine for the selection of optimal smoothing parameters and the corresponding heat maps accurately identify the unique patterns in which players hit the baseball on the baseball field.

1 Introduction

Statistical data science plays a significant role in baseball with large amounts of data being collected during every game for the purpose of analysis, evaluation and prediction of players' performance. Batting average and on-base percentage are some examples of the statistics that are used to evaluate the offensive performance of a hitter whereas earned run average and the number of strikeouts are some examples of the statistics that help explain the performance of a pitcher. However, these statistics are not sufficient to fully understand a player's performance. By using spatial statistics methodology and exploratory tools we can improve the analysis of players' performance by including additional information about players' skills. For instance, heat maps provide better insights about players' strengths and weaknesses. Cross and Sylvan (2015) [3] used a spatial covariance function to model spatial batting ability and developed a method for producing more accurate heat maps than the traditional ones by using kriging (spatial interpolation), a techniques common to Geostatistics. The resulting heat maps revealed individual characteristics of a hitter based on the locations of pitches inside and outside the strike zone. Heat maps are used not just for hitters but also for pitchers. Wilcox and Mannshardt (2013) [7] developed a model for the intensity function for pitchers based on the location of pitches and identified the spatial pattern of these pitches from the resulting heat maps.

The objective of this paper is to analyze the point pattern data of batted balls for several top Major League Baseball players and understand their offensive performance through the analysis of heat maps. A detailed description of the data is provided in next section, including creation of new spatial variables with corresponding R references. Exploratory visuals are shown here as well. A summary of the methodology for spatial point patterns is given in Section 3 and is illustrated in Section 4 where we show heat maps of batted balls for the selected players. Concluding remarks and some future directions are given in last section.

2 Data

We use data for the three players who won the Silver Slugger Award between 2018 and 2022. They are Freddie Freeman, Paul Goldschmidt, and Josh Bell. This award is given to the best offensive players at their positions. They were selected by league managers and coaches, see Casella (2022) [2]. We chose not to use data from the 2020 season since it was a shortened season and the sample size was significantly smaller compared to other seasons. The function **scrape_statcast_savant_batter()** from the R **baseballr** package was employed to extract the data for each of these players from `baseballsavant.mlb.com` by entering the corresponding batter ID, the start date and end date of each of 2018, 2019, 2021 and 2022 seasons, respectively. After some data cleaning, we extracted the variables *hc_x* and *hc_y* which are the x- and y-coordinates of the batted balls for each at-bat from the data set of each player. We then created new variables *hc_x_adjusted* and *hc_y_adjusted*, which are the modified values of *hc_x* and *hc_y* by using the formulas discussed by Petti (2017) [6] as follows.

$$hc\_x\_adjusted = hc\_x - 125.42 \tag{1}$$

$$hc\_y\_adjusted = 198.27 - hc\_y \tag{2}$$

These adjustments flip the points around and make the origin $(0,0)$ the home plate of the baseball field. This simplifies the R code used to produce the point pattern data set and the subsequent heat maps. Figure 1 shows scatterplots of the batted balls for Paul Goldschmidt, Freddie Freeman and Josh Bell, respectively. Goldschmidt is a right-handed hitter, Freeman is a left-handed hitter and Bell is a switch hitter, who can hit from both sides of the batter's box.

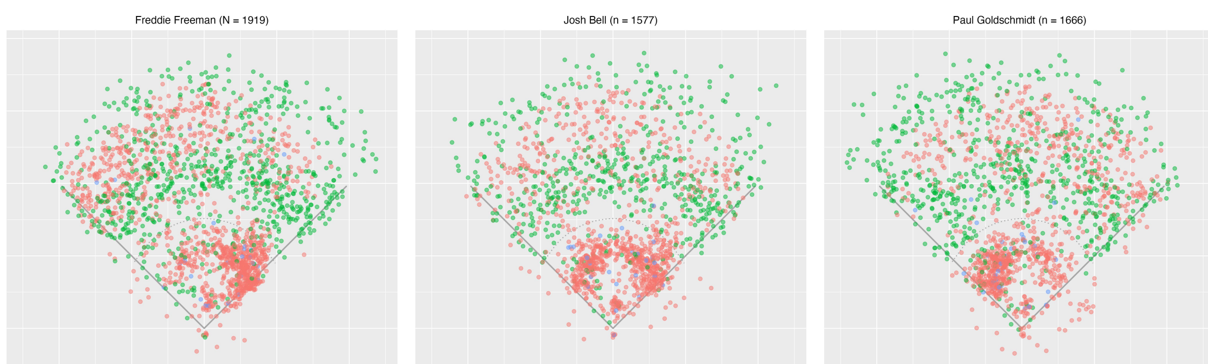


Figure 1: Scatterplots of batted balls for Freddie Freeman (left), Josh Bell (center) and Paul Goldschmidt (right); green dots mark hits, red dots mark outs, blue dots mark error/fielder's choice

The x-axis shows the x-coordinates of the batted balls and the y-axis shows the y-coordinates of the batted balls. The V-shaped lines represent the foul lines of the baseball field. The different colors mark different outcomes for the batted balls. Based on these scatterplots, we can see a cluster of points on the left side of the infield of the baseball field for Goldschmidt. In contrast, there is a cluster of points on the right side of the infield of the baseball field for Freeman. As for Bell, there is a cluster of points on both sides of the infield of the baseball field.

In order to analyze how the players hit, we create a new point pattern data set for each player by using the R function **ppp()** from the **Spatstat** package. This function requires the user to enter x- and y-coordinates of the batted balls for each at-bat as well as the ranges of the x-coordinates and y-coordinates. In order for the point pattern data set to contain all the points of the batted balls from the original data set of the players, the range of the x-coordinates was determined based on the leftmost and rightmost points in the data set of each player for the aggregated seasons 2018–2022. Similarly, the range of the y-coordinates was determined based on the highest and lowest points in the data set of each player for the last four seasons. For example, in the data set for Freddie Freeman, the x-coordinates of the leftmost point and rightmost point are, -109.48 and 118.50 and therefore, the range of the x-coordinates is $[-110, 119]$. Likewise, the y-coordinates of the highest and lowest points are 188.23 and -21.77, respectively, and thus the range of the y-coordinates is $[-22, 189]$. As a result of choosing the values for the ranges of x- and y-coordinates, the dimension of the point pattern data set for Freddie Freeman is 229 units by 211 units. These point pattern data sets for the players are further used for chi-squared homogeneity tests, smoothed heat maps, and visualization.

3 Methods

We first performed chi-squared tests for spatial homogeneity for each of the three players. In all instances the hypothesis of complete spatial randomness was rejected. This implies that the intensity function of batted balls varies with the space location. The underlying stochastic process can therefore be viewed as an inhomogeneous Poison process or point pattern. For each player the first order intensity function of batted balls is defined as

$$\lambda(s) = \lim_{|ds|\to 0} \frac{E[N(ds)]}{|ds|} \tag{3}$$

for any point $s \in A$, where A is a random domain and $N(A)$ is the random variable counting the number of events in A. The connection with the average number of events in region A is given by

$$E[N(A)] = \int_A \lambda(s)ds. \tag{4}$$

Studying point patterns through $\lambda(s)$ rather than through $E[N(A)]$ is mathematically advantageous because it eliminates the dependency of the size and shape of the area A. However, context is important in practical applications, the illustrations that we show in next section are based on the V-shaped domain previously described.

Next we estimate the first order intensity function nonparametrically by using kernel smoothing as

$$\hat{\lambda}(s_0) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{s_i - s_0}{h}\Big). \tag{5}$$

Here K is a smooth, known function (the kernel) and h is the smoothing parameter (bandwidth). We use the method proposed in Diggle (1985) [4], which is implemented in R and contains a routine for optimal bandwidth selection. The function **bw.diggle()** in the **spatstat** package calculates the bandwidth that minimizes the mean squared error (MSE). The algorithm in this function utilizes the method of Berman and Diggle (1989) [1] with the optimal bandwidth h chosen so as to minimize

$$M(h) = \frac{MSE(h)}{\lambda^2} - g(0) \tag{6}$$

with respect to the smoothing parameter h. Here g denotes the correlation function. For a comprehensive description of kernel estimation of intensity functions for inhomogeneous Poisson processes we refer to the textbook Diggle (2003) [5].

4 Results

To illustrate the methodology previously described we show below the heat maps indicating spatial patterns of batted balls for the selected players. The V-shaped lines represent the foul lines of the baseball field which will help us locate the cluster of points in the corresponding area on the actual baseball field. The estimated point patterns for Freeman, Bell and Goldschmidt are shown in Figure 2.

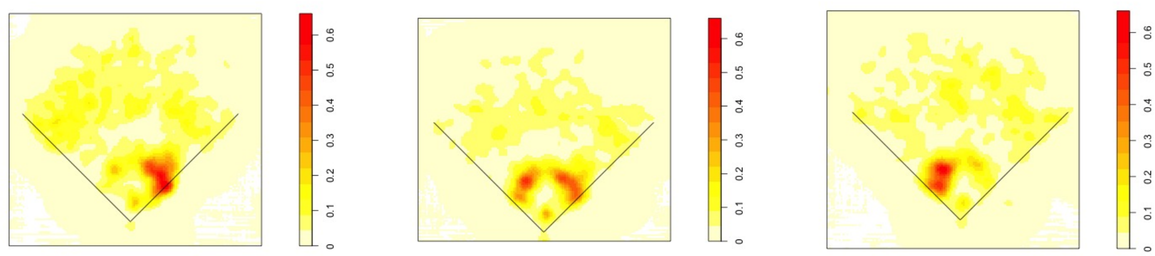


Figure 2: Heat maps for Freddie Freeman (left), Josh Bell (center) and Paul Goldschmidt (right)

Based on the heat map for Freeman (left), the orange and red areas on the right side of the infield of the baseball field in the maps indicate that he tended to hit towards the right side of the infield of the baseball field. As for Bell (center), the red and orange areas on both sides of the infield of the baseball field in the maps indicate that there are clusters of points on the both side of the infield of the baseball field. This implies that Bell hit evenly to both sides of the infield of the baseball field. In the heat maps for Goldschmidt (right), the location of the red and oranges areas imply that Goldschmidt tended to hit to the left side of the infield of the baseball field.

5 Discussion

The analysis of the point pattern data set of the batted balls help understand the hitters' performance and abilities. Rejection of the complete spatial randomness hypothesis indicated the presence of clusters for batted balls. The heat maps helped accurately identify where these clusters are on the baseball field. Freeman, who is a left-handed hitter tends to hit to the right side of the infield of the baseball field whereas Goldschmidt, who is a right-handed hitter, tends to hit to the left side of the infield of the baseball field. There were clusters of points on both sides of the infield of the baseball field for the point pattern data for Bell who is a switch hitter. In future work we will consider the analysis of the batted balls more in depth by splitting the data set by the pitches they hit or by the types of pitchers they faced (for non-switch hitters). In addition, we will update our analysis by using data for batted balls from future seasons. It will also be interesting to determine whether the new rule which bans the use of defensive shifts will affect the way players hit baseballs by comparing them with the batted balls from the past seasons.

References

[1] Berman, M. and Diggle, P.J. (1989). Estimating Weighted Integrals of the Second-Order Intensity of a Spatial Point Process, Journal of the Royal Statistical Society (B), 51(1), 81–92.

[2] Casella, P. (2022). Here are the 2022 Silver Slugger winners. (online)

[3] Cross, J. and Sylvan, D. (2015). Modeling spatial batting ability using a known covariance matrix, Journal of Quantitative Analysis in Sports, 11(3), 155-167.

[4] Diggle, P. J. (1985). A Kernel Method for Smoothing Point Process Data. *Applied Statistics,* 34(2), 138-147.

[5] Diggle, P.J. (2003) *Statistical analysis of spatial point patterns*, 2nd ed. Arnold.

[6] Petti, B. (2017) The Hardball Times. (online)

[7] Wilcox, A. and Mannshardt, E. (2013). Baseball Scouting Reports via a Marked Point Process for Pitch Types (online)

Defensive team analysis in the 2022 World Cup based on event prediction

Rikuhei Umemoto[1], Kazushi Tsutsui[1], Keisuke Fujii[1,2,3,*]

[1] Graduate School of Informatics, Nagoya University, Nagoya, Aichi, Japan.
[2] RIKEN Center for Advanced Intelligence Project, Fukuoka, Fukuoka, Japan.
[3] PRESTO, Japan Science and Technology Agency, Kawaguchi, Saitama, Japan.
* fujii@i.nagoya-u.ac.jp

Abstract

Analyzing defenses in team sports is challenging because of the limited event data. Researchers have previously proposed methods to evaluate football team defense by predicting the events of ball gain and being attacked. However, they did not consider the importance of the events, and did not fully investigate the influence of the diversity (e.g., nationality). Here, we propose a generalized valuation method of defensive teams by score-scaling the predicted probabilities of the events. Using the location data of all players in football games of the 2022 World Cup, we investigated the effect of the number of players on the prediction and validated our approach by analyzing the games. Results show that for the predictions of being attacked, scoring, and conceding, all players' information was not necessary, while that of ball gain required information on three to four offensive and defensive players. We also explained the excellence in defensive teams in the 2022 World Cup.

1 Introduction

Football has one of the largest number of players and spectators in the world. Due to its diversity, it is necessary to establish a common metric to quantitatively evaluate a football team, regardless of the background of the players or spectators. However, the analysis of football data is usually difficult due to the continuous movement of the ball and players during a match and the tendency to be limited to specific event data (e.g., scores and concedes). In particular, defensive tactics are considered difficult to evaluate due to the limited statistical information (for an overview, see review [3]). Recently, tracking techniques and football videos have been made publicly available (see e.g., [5, 1]), but a large amount of accurate whole-time tracking data for all players remains unpublished. Therefore, the evaluation of attacking players with the ball is often based on event data and the ball coordinates. For example, previous studies evaluated play by the expected value of scores and concedes calculated by their probabilities using location data of the ball (e.g., [2, 10]).

Studies on the evaluation of off-ball players (e.g., in the review of [6, 7]) required location data for all players, but these are usually private. Recently, StatsBomb Inc. (football analytics and data visualization company in U.K.) has published all players' location data in broadcast video frames for soccer games of men's Euro 2020, women's Euro 2022, and FIFA World Cup 2022 football competitions. However, there is little research on the evaluation of defensive play. For team evaluation using tracking data of all players, researchers proposed a method to evaluate team defense by predicting ball gains (or ball recoveries) and effective attacks[8]. However, this method has the following problems: (1) the weight in the formula did not represent the importance, (2) the method assumed complete location data for all 22 players and the impact of the lack of data for all players is unknown, and (3) this study only was conducted in a domestic men's professional league and diversity (such as country) has not been fully investigated.

To solve these problems, we first propose a generalized method of valuing defensive teams by scaling the predicted probabilities of scores and concedes, called Generalised Valuing Defence by Estimating Probabilities (GVDEP). We then investigate the impact of reducing the number of players on the prediction model using all players' locations in a broadcast video frame in FIFA World Cup 2022 football matches Lastly, we validate our proposed method for analyzing the defense of the top 16 teams.

2 Methods

2.1 Dataset

We used the open-source dataset of FIFA World Cup games held in 2022 provided by StatsBomb Inc. (UK) because it contains the location data of players of the teams in diverse nations. The Statsbomb dataset has been widely used in academic studies (e.g., [4, 2]). It would be guaranteed that the use of the data would not infringe on any rights of players or teams. The dataset is available at `https://github.com/statsbomb/open-data`.

Also, the dataset includes event data (i.e., labels of actions and xy coordinates of the ball). This dataset contains 120,278 event data, of which 152 out of 1,277 shots were successful, 7,167 effective attacks were played, and 2,395 ball gains were realized. An effective attack was defined as an event that ends in shooting or penetrating into the penalty area, and a ball gain was a change in the attacking by some factors. In this study, an effective attack is defined as *attacked* from the viewpoints of defenders. It should be noted that we labeled for each event (not an attack segment) whether positive/negative ball gains or attacked occurred at an event.

In addition, the dataset has location data (i.e., xy coordinates) of players in any broadcast video frame. Many scenes of a soccer broadcast video do not show us all 22 players. Therefore, note that some in this dataset do not include all 22 players' information. In the case of this dataset, the minimum, first quartile, median, third quartile, and maximum values of the number of players in each frame were 0, 11, 15, 17, and 22 respectively.

2.2 Proposed Method: GVDEP

We illustrate our proposed method, GVDEP, for evaluating a defending team. Suppose that the ith state of a game is given by s_i and the set of it is given by $S = [s_1, \ldots, s_N]$ in chronological order. This s_i includes the event or action features a_i, the on-ball features b_i and the off-ball features o_i. Note that we show the result of the validation of the number of players in the following section. To evaluate all state transitions for defensive and offensive actions in this study from the defender's perspective, the following time index i is used as the ith event. We define the probability of future ball gain $P_{gains}(s_i)$ and attacked $P_{attacked}(s_i)$ in the game state s_i of a certain interval at an ith event. In the previous study[8], these probabilities were used to propose VDEP expressed as $V_{vdep}(s_i) = P_{gains}(s_i) - C * P_{attacked}(s_i)$, where C is the frequency ratio of ball gains and attacked in the training data. However, this is not adjusted based on the importance. Hence, in this study, we propose GVDEP (V_{gvdep}) weighted on the predicted probabilities of scores and concedes used in VAEP[2] as follows:

$$V_{gvdep}(s_i) = w_{gains} \times \Delta P_{gains}(s_i) - w_{attacked} \times \Delta P_{attacked}(s_i), \tag{1}$$

$$w_{gains} = \sum_{j \in Ev_{gains}} \frac{sign(Teams_j) V_{vaep}(s_j)}{|Ev_{gains}|}, \tag{2}$$

$$w_{attacked} = \sum_{j \in Ev_{attacked}} \frac{sign(Teams_j) V_{vaep}(s_j)}{|Ev_{attacked}|}, \tag{3}$$

where Ev_{gains} and $Ev_{attacked}$ are the set of ball gains and attacked, $|Ev_{gains}|$ and $|Ev_{attacked}|$ are the total number of each set. Also, $sign(Teams_j)$ is the sign indicating which team caused a ball gain or attacked. Furthermore, we calculated P_{scores} and $P_{concedes}(s_i)$, which are the probability of scores or concedes in the future respectively, for using the framework of VAEP[2] in our formula. In addition, to evaluate the team defense, we defined $R_{gvdep}(p)$ as the evaluation value per team p as follows: $R_{gvdep}(p) = \frac{1}{M}\Sigma_{s_i \in \boldsymbol{S}_M^p} V_{gvdep}(s_i)$, where M is the number of events for team p defending in all matches, and $\boldsymbol{S}_M^p$ is the set of states S of team p defending in all matches. The code is available at `https://github.com/Rikuhei-ynwa/Generalized-VDEP`.

2.3 Predicting the Probabilities

Regarding the previous studies[2, 8], the set of features s_i to be input to the classifiers used in this study was constructed using the event tracking data provided by StatsBomb. First, we used the event type (24 types such as pass and shot) as event a_i. Yet, a_i was used in the classifiers for VAEP but not used for VDEP due to the differences in their respective approaches. Next, as on-ball features b_i, we used the start/end point and time of the event, the movement displacement and elapsed time from the start to the end of the event, the distance and angle between the ball and the goal, and whether there was an offensive or defensive change from the previous event (21 dimensions in total). In this study, the off-ball feature o_i at the time of the event was also included in the state s_i. Specifically, for each team, we used data in which the xy coordinates representing the respective positions of all players and the distance of each player to the ball were sorted in order of proximity to the ball (88 dimensions in total).

The classifiers for predicting ball gains, attacked, scores, and concedes were the XGBoost (eXtreme Gradient Boosting) classifier used in the previous studies. Gradient boosting methods are known to perform well on a variety of learning problems with inhomogeneous features, noisy data, and complex dependencies. According to the previous studies[2, 8], we used $s = [a, b, o]$ for the ith and $i-1$th actions as the features.

For the classification problem of estimating $P_{gains}(s_i)$, we assign a positive label ($= 1$) to the game state s_i if a defending team in state s_i was able to gain the ball in the subsequent k event, and a negative label ($= 0$) if it did not gain the ball. Similarly, for the classification problem of estimating $P_{attacked}(s_i)$, we assign a positive label ($= 1$) to the game state s_i if a defending team caused attacked in the subsequent k event (the same for scores and concedes). The k is a parameter freely determined by the user; a small k gives a short-term reliable prediction, while a large k gives a long-term prediction including many factors. In this study, following previous studies, $k = 5$[8] is used for predicting P_{gains} and $P_{attacked}$, and $k = 10$[2] is used for predicting P_{scores} and $P_{concedes}(s_i)$. Also, 15,498 of the 120,278 events in this dataset had no data on the xy coordinates of all players. Therefore, 104,780 events with that data removed were used in the classifier. This resulted in 1,274 scores, 242 concedes, 4,229 ball gains, and 17,479 attacked out of the number of events defined by k above. From these results, as mentioned in the previous study[8], it can be interpreted that scores and concedes are rarer events than ball gains and attacked. Therefore, in this study, scores and concedes are examined with a smaller dataset (compared to the large one of the previous study[2]).

2.4 Validation, Evaluation and Analysis

First, the classifiers for predicting the probabilities were validated. In the evaluation of intuitively understandable accuracy and so on, there is a possibility that the evaluation cannot be done correctly when the number of negative examples is extremely larger than positive ones, as in this study and in the previous study[8]. Therefore, in this study, we used the F1 score, which is one of the methods to evaluate whether true positives can be classified without considering true negatives. The F1score is calculated using Precision, which indicates the proportion of items classified as positive that are true positives, and Recall, which indicates the proportion of items that should have been classified as positives but were correctly classified as positives.

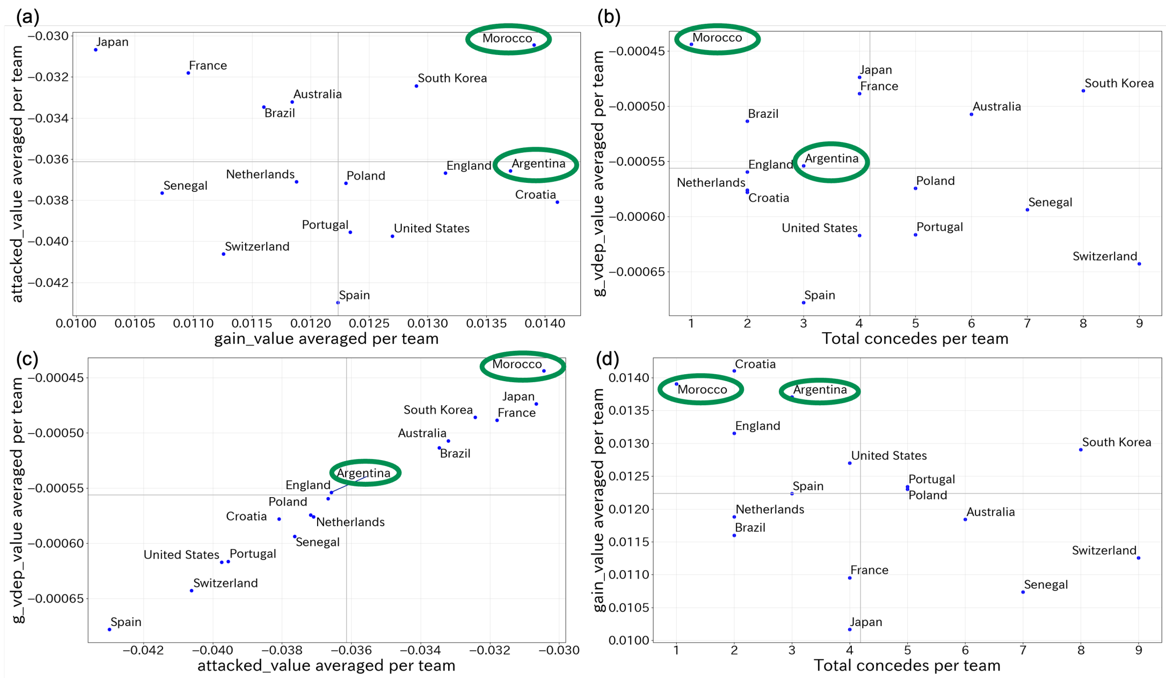


Figure 1: **Defense evaluation of top 16 teams in FIFA World Cup 2022.** Diagram showing the relationship between (a) *gain_value* and *attacked_value*, (b) concedes and GVDEP, (c) *attacked_value* and GVDEP, (d) concedes and *gain_value* in FIFA World Cup 2022. Here, green circles represents the teams (Morocco and Argentina) used as examples in this paper.

Also, to validate the performance of the classifiers, it is necessary to separate the training data from the test data. In this study, the number of matches for training and validation was divided into $58:6$ or $57:7$ in the manner of cross-validation, and a total of 10 cross validations were carried out, six and four respectively. In contrast, in the final football analysis, we predicted the probabilities using all matches for training and testing to prioritize the use of all data over the validation of generalization performance. In other words, to calculate the GVDEP, we trained the classifiers using all 64 matches (FIFA World Cup 2022) and predicted ones using the same ones. Finally, the probabilities and equations obtained from these predictions were used to calculate GVDEP. During team defense analysis, we analyzed the top 16 teams using 48 group stage matches and 8 round 16 matches. Here, Pearson's correlation coefficient r is used for the correlation analysis between variables, and $p < 0.05$ is considered significant in all statistical analyses.

3 Results

As a validation of GVDEP, we first examined the prediction performance when the number of attackers/defenders around the ball (*n_nearest*) is varied. Here, the probabilities of ball gains (P_{gains}), attacked ($P_{attacked}$), scores (P_{scores}), and concedes ($P_{concedes}$) are predicted using different classifiers. The tendency was similar to those of Euro 2020 and 2022 [9]. In summary, the F1-score of P_{gains} improved for values of n_nearest up to 3 and 4, while it did not improve for later values of n_nearest. On the other hand, the F1-scores of P_{scores}, $P_{concedes}$, and $P_{attacked}$ did not improve for all players' location data.

Next, examples of defense evaluation of the top 16 teams in FIFA World Cup 2022 are shown in Figure 1. In Figure 1, the average values of ΔP_{gains} and $\Delta P_{attacked}$ were defined as *gain_value* and *attacked_value* respectively, and R_{gvdep} was as *g_vdep_value*. First, we characterize and evaluate team defense using the *gain_value* and *attacked_value* in Figure 1(a). We couldn't identify a trade-off between the *gain_value* and *attacked_value* averaged per team ($r = -0.118, p = 0.664$). This result is not similar to the previous study[8]. Next, we showed the relationship between the actual number of concedes and *g_vdep_value* averaged per team in Figure 1(b), but we couldn't find a significant correlation between the two ($r = -0.197, p = 0.464$). As this approach evaluates the process of defense, the difference between the two is more important than the correlation itself. However, it should be noted that there was a strong correlation ($r = 0.993, p = 3.089 \times 10^{-14}$) between *g_vdep_value* and *attacked_value* in Figure 1(c). This may be because the fact that the absolute value of $|weight\_attacked|$ (= 0.021) was larger than the value of $|weight\_gains|$ (= 0.011). Thus, it should be noted that the interpretation of the results in GVDEP is similar to that of *attacked_value*. Finally, for this reason, we examined the relationship between the actual number of concedes and *gain_value* in Figure 1(d) and found no significant correlation between them either ($r = -0.432, p = 0.0951$). However, roughly speaking, the figure shows that teams are less at risk of concedes if they try to get themselves into a position to gain the ball in the future. This result is different from the case of UEFA EURO 2020 analyzed in the previous study[9].

4 Discussion

In this study, we first validated the existing probabilities based on the prediction performance. To calculate the probabilities of ball gains (P_{gains}) and attacked ($P_{attacked}$), the previous study [8] used data including all player's features. However, the type of this data is not always available because this is often private or expensive. Hence, we used open-source data including all player's location data in a video frame, and we validated the existing classifiers' performances. This result suggests that although features of not only the ball but the players are important to improve the classifier in ball gain, we do not need all players' locations. Our results suggest that our approach can evaluate defensive performances only with open-source data.

Next, we quantitatively analyzed the games in FIFA World Cup 2022 using our proposed method. According to the correlation analysis, we found the trade-off between the tendencies of ball gains and of attacked. However, Morocco was able to maintain a good balance between both the ball gain and not being attacked at a high level. Therefore, they acquired the highest *g_vdep_value* in the top 16 teams. This suggested that their defense may have been an organized and effective defense that focused on nullifying opposition attacks and preventing scoring opportunities. This could explain why they reached the top four in the World Cup for the first time in African history. Also, the World Cup winner, Argentina, took a higher *gain_value* and almost average *attacked_value*. This suggested that they may have had an average transition to being able to gain the ball. In fact, Argentina attempted more tackles and interceptions than the average of the top 16 teams.

Finally, we introduce the limitations of this study and future perspectives. The first issue is that our formula is too affected by *attacked_value*. As shown in Figure 1(c), we found a too-high positive correlation between *attacked_value* and GVDEP values. It is true that not allowing the opponent to effectively attack can be seen as reducing the probability of conceding a goal, but it is difficult to evaluate the defense by only this feature. The second problem is that the proposed indicator is not able to evaluate any given team. The second is the definition of off-ball features and the modeling. In this study, the features only include 22 players' *xy* coordinates, the distance, and the angle of each player from the ball.

In conclusion, we proposed a method to evaluate team defense. In the validation, we investigated the impact of reducing the number of players on the prediction model and analyzed the defense of the teams that reached round 16 in the World Cup 2022. The result indicated that their decision-making skill is important as well as a runner's sprint speed and base running skill. For future work, we can consider using more data, more specific features of the off-ball defense, and another nonlinear modeling such as using neural networks.

Acknowledgments

This work was supported by JSPS KAKENHI (Grant Number 20H04075) and JST Presto (Grant Number JPMJPR20CA).

References

[1] A. Cioppa, S. Giancola, A. Deliège, L. Kang, X. Zhou, Z. Cheng, B. Ghanem, and M. Van Droogenbroeck. Soccernet-tracking: Multiple object tracking dataset and benchmark in soccer videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 3491–3502, June 2022.

[2] T. Decroos, L. Bransen, J. Van Haaren, and J. Davis. Actions speak louder than goals: Valuing player actions in soccer. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1851–1861, 2019.

[3] K. Fujii. Data-driven analysis for understanding team sports behaviors. *Journal of Robotics and Mechatronics*, 33(3):505–514, 2021.

[4] S. Gregory, S. Robertson, R. Aughey, and G. Duthie. The influence of tactical and match context on player movement in football. *Journal of Sports Sciences*, 40(9):1063–1077, 2022.

[5] A. Scott, I. Uchida, M. Onishi, Y. Kameda, K. Fukui, and K. Fujii. Soccertrack: A dataset and tracking algorithm for soccer with fish-eye and drone videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3569–3579, 2022.

[6] W. Spearman. Beyond expected goals. In *Proceedings of the 12th MIT Sloan Sports Analytics Conference*, pages 1–17, 2018.

[7] M. Teranishi, K. Tsutsui, K. Takeda, and K. Fujii. Evaluation of creating scoring opportunities for teammates in soccer via trajectory prediction. In *International Workshop on Machine Learning and Data Mining for Sports Analytics*. Springer, 2022.

[8] K. Toda, M. Teranishi, K. Kushiro, and K. Fujii. Evaluation of soccer team defense based on prediction models of ball recovery and being attacked: A pilot study. *Plos one*, 17(1):e0263051, 2022.

[9] R. Umemoto, K. Tsutsui, and K. Fujii. Location analysis of players in uefa euro 2020 and 2022 using generalized valuation of defense by estimating probabilities. *arXiv preprint arXiv:2212.00021*, 2022.

[10] C. C. K. Yeung, T. Sit, and K. Fujii. Transformer-based neural marked spatio temporal point process model for football match events analysis. *arXiv preprint arXiv:2302.09276*, 2022.

Causal problems involving football strategy

Tim B. Swartz*

* Dept of Stats/ActSci, Simon Fraser University, Burnaby BC, Canada V5A1S6 email: tswartz@sfu.ca

Abstract

This paper investigates three problems involving strategy in football: (a) whether to cross the ball, (b) whether teams should play with pace and (c) what are optimal throw-in locations. The problems are addressed using tracking data which facilitates the identification and quantification of confounding variables that are related to the outcome variables and the treatment variables. Although the data structures differ in the three problems, the line of research provides a template for addressing causal inference in sport. From a practical viewpoint, we obtain some surprising results.

1 Introduction

A primary problem in sport is the application of strategy. However, the determination of whether a strategy is beneficial is a challenging problem. A reason for this is that cause and effect relationships in sport cannot typically be investigated via randomized experiments. Sporting events yield observational data which at best can only determine associations between strategies and outcomes.

Fortunately, the development of causal methods (Pearl 2009) provides opportunities to investigate cause and effect relationships in non-experimental settings. Whereas the identification and measurement of relevant confounding variables is a necessary and challenging component of causal methods, the hurdle appears less imposing in sport. In sport, objectives are often clear (e.g. score more goals than the opponent), matches terminate in reasonable timeframes (e.g. often two to four hours), and rules are well defined. Most importantly, with the advent of detailed player tracking data (e.g. spatio-temporal data), our sporting intuition often permits the identification and measurement of relevant confounding variables.

With football tracking data, the location coordinates for every player on the field are recorded frequently (e.g. 10 times per second). With such detailed data, the opportunity to explore novel questions in sport has never been greater. The massive datasets associated with player tracking also introduce data management issues and the need to develop modern data science methods beyond traditional statistical analyses. Gudmundsson and Horton (2017) provide a review of spatio-temporal analyses that have been used in invasion sports where player tracking data are available.

In Section 2, we provide a template for addressing problems of strategy in sport under a causal methods framework with the assistance of tracking data. In Section 3, these methods are applied to three problems in football: (a) whether to cross the ball, (b) whether teams should play with pace and (c) what are optimal throw-in locations. Some surprising results are obtained which are of practical value. We finish with some concluding remarks in Section 4.

2 Methods

In what follows, we develop a general template for the investigation of cause and effect relationships in sport given the availability of tracking data.

The common structure in the football problems that we consider involves the cause and effect relationship between a variable X (a strategy) on an outcome variable Y. In these problems, we also introduce a multivariate confounding variable W where W impacts both X and Y. In terms of methods, our problems differ in the nature of the variables (W,X,Y). For example, in some problems X may be binary whereas in other problems, X may be multivariate and continuous.

In the following problems, football intuition and the availability of tracking data allow us to define and collect data W for each instance of (X,Y). It is the identification and measurement of W which is necessary to apply the causal methods described here. For example, let's suppose that an attacker is considering the strategy $X = 1$ to deke a defender versus the strategy $X = 0$ to shoot. We may consider defining an outcome variable $Y = 1$ corresponding to an imminent goal and $Y = 0$ corresponding to not scoring a goal. In this situation, there are many considerations (i.e. W) that affect both X and Y. For example, if the attacker is moving fast relative to the primary defender, it may increase the chance of deking versus shooting. Moving fast may also increase the chance of scoring.

In the above setup, a statistical model is then proposed which is fit and permits the calculation of propensity scores $\text{Prob}(X \mid W)$. Now suppose that the strategy variable X takes on k distinct strategies or is discretized to take on k distinct strategies where each strategy has been observed a minimum of n times in the dataset. Consider then a strategy S which has n observations. A matching algorithm is utilized (Austin 2011) where each observation in S is matched with an observation from the other $k-1$ strategies such that the pairs of strategies have similar propensity scores. Consequently, we have k groups (i.e. strategies) each with n observations which are balanced in terms of propensity scores. It is this balance which mimics the randomization of treaments to subjects. We then measure the effectiveness of the k strategies X by observing their corresponding outcome variable Y. This may be done inferentially or via descriptive statistics.

3 Causal Problems in Football

In this section, we describe three problems related to football strategy. The analysis is based on tracking data from 237 matches from the 2019 season of the Chinese Super League.

3.1 Crossing the Ball

For greater detail on the crossing problem, the reader is referred to Wu et al. (2021).

Here, we are interested in whether an attacker along the wing in the final third of the field should cross the ball ($X = 1$) or not cross the ball ($X = 0$). Although success is typically measured by goals, goals are rare events, and we therefore define the outcome variable according to whether a shot was generated ($Y = 1$) or not ($Y = 0$) by the end of the possession. An example of a measurable confounding variable W in this problem is the number of attackers relative to defenders in the box. With more attackers, the crosser is more likely to cross and a shot is more likely to occur.

With a propensity score model described in Wu et al. (2021), the methods of Section 2 are implemented based on $k = 2$ strategies. The inferential component of our investigation begins with the simple two-sample test of proportions between the two groups (i.e. strategies) based on the response Y (resultant shot). The quantity of interest is the average treatment effect $\text{ATE} = \bar{Y}(1) - \bar{Y}(0)$ where $\bar{Y}(1)$ is the mean number of resultant shots when the ball is crossed and $\bar{Y}(0)$ is the mean number of resultant shots when the ball is not crossed. We obtain $\text{ATE} = 0.050$ with standard error 0.020. The result is significant and suggests that crossing is beneficial in the sense that a cross will lead to a shot 5% more often than when not crossing.

In Figure 1, we smooth the variable Y for each strategy with respect to the propensity score. We observe that as the propensity score increases (i.e. conditions become more favourable to crossing) the probability of a resultant shot increases for both groups. However, and most importantly, we observe that the shot probability of the treatment group increases relative to the shot probability of the control group. In practice, what this means is that players are making correct tactical decisions. When players are more likely to cross (higher propensity scores), they will have better offensive results (higher shot probabilities) than if they did not cross.

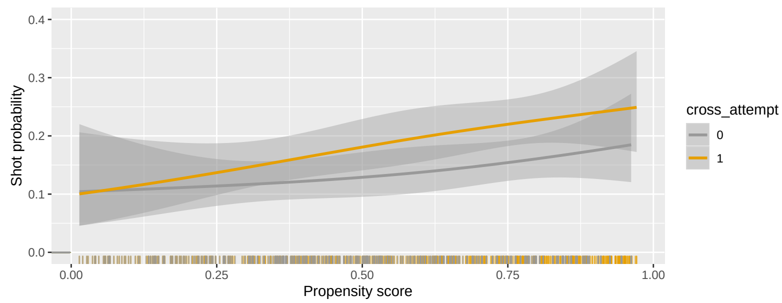


Figure 1: After matching, smoothed plots of the shot variable Y for both groups (crossing versus non-crossing) with respect to the propensity score.

Therefore, the takeaway message is that crossing (when done under the right circumstances) is a good thing to do. And players do cross at the right times.

3.2 Playing with Pace

For greater detail on the pace of play problem, the reader is referred to Epasinghege Dona and Swartz (2023a).

Here, we are interested in the adage that playing fast is better than playing slow (Blank 2012). Whereas this is an appealing notion, there is no univerally accepted definition of playing with pace, and the adage has not been tested against data.

We introduce a definition of attacking pace that concerns movement towards the attacking goal. Thus, the tiki-taka style popularized by Barcelona (2005-2011) is not an example of playing with pace according to our definition. Accordingly, we have a continous variable X which is the pace differential between the home and road teams over a 15-minute time interval. In this case, the outcome variable Y is the excess shots by the home team. An example of a measurable confounding variable W is the goal differential. As a team's lead increases they tend to play more conservatively with less attacking pace.

With the above setup, we discretize X into $k = 2$ groups according to whether the home (road) team is playing with greater pace. A statistical model is then proposed which permits the calculation of the propensity scores $\text{Prob}(X \mid W)$. In Figure 2, we smooth the variable Y with respect to the propensity score for each group. We observe that as the propensity score increases (i.e. conditions become more favourable for the home team to play at greater pace), the excess shots for the home team increases for both groups. We also observe that the excess shots by the home team remains relatively constant across the two groups as the propensity score increases. In practice, this means that the advantage of playing at pace persists no matter the circumstances that dictate whether a team should play at pace. Over the range of propensity scores $\text{Prob}(X \mid W)$, it seems that the home team generates approximately an extra 1/2 shot compared to the road team on average during a 15-minute interval.

Therefore, the takeaway message is that playing with pace is a good strategy. It leads to more shots for than against. This provides support to Blank's thesis - the Holy Grail of tactics (in Chapter 1 of Blank (2012)) that fast is better than slow.

3.3 Throw-ins

For greater detail on the throw-in problem, the reader is referred to Epasinghege Dona and Swartz (2023b).

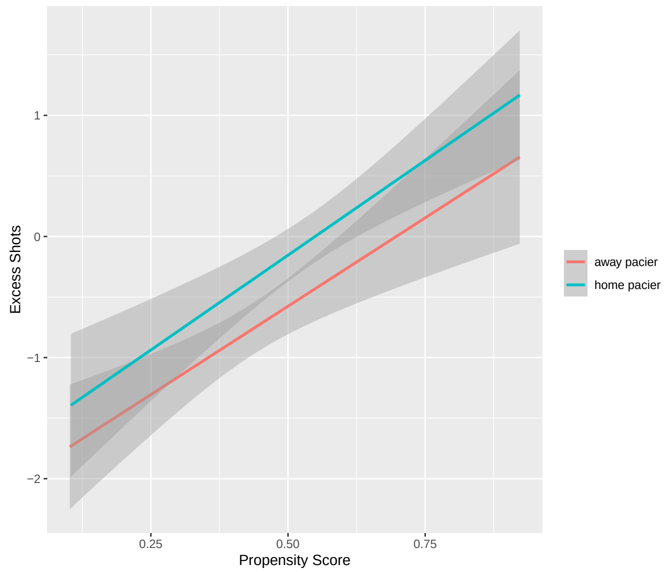


Figure 2: After matching, smoothed plots of the excess shot variable Y for the home team with respect to the propensity score under the treatment (blue) and the control (red).

Here, we are interested in the spatial location X where a throw-in is received. Is there a strategic element in choosing X? The outcome variable is binary where a shot was generated ($Y = 1$) or was not generated ($Y = 0$) by the end of the possession. An example of a confounding variable W is the location where the throw-in takes place. As the location moves further in the attacking end, there is less room to make a forward throw-in, and being closer to goal, there is a greater chance for a resulting shot.

In this problem, we discretize the spatial location X into $k = 200$ regions each of 4.0 squared metres. Again, a statistical model is proposed and fit which allows the calculation of propensity scores $\text{Prob}(X \mid W)$.

To investigate the causal effect of X on Y, we produce a smoothed heat map of $\bar{Y}$ in Figure 3 that has been smoothed using the function *interp.loess* in the R package *tgp*. We observe darker regions (i.e. larger $\bar{Y}$) to the left (backward throw-ins) and near the top (long throw-ins). Whereas long throw-ins seem more common today than in the past, there remain more forward throw-ins than backward throw-ins.

4 Concluding Remarks

Given the availability of tracking data, we have provided a template for the investigation of strategy in sport using causal methods. This relies on sporting intuition such that confounding variables W that affect both the strategy X and the outcome Y can be identified and measured. We hope that this work will motivate future causal investigations regarding strategy in sport.

On a practical note, we have investigated three causal problems in football which yield noteworthy results. First, although there appears to be a decreasing reliance on crossing the ball in top-level football, our investigation suggestions that crossing the ball remains a valuable tactic. Moreover, it appears that players know when it is best to cross the ball and when not to cross the ball. Second, moving the ball quickly in an attacking direction is seen as a valuable tactic. Although there is some support for this notion, it has not previously been validated via data. Third, our analysis provides support for both long throw-ins and backwards throw-ins. Although long throw-ins are seemingly on the rise, there currently does not appear to be much chatter regarding the benefit of backwards throw-ins.

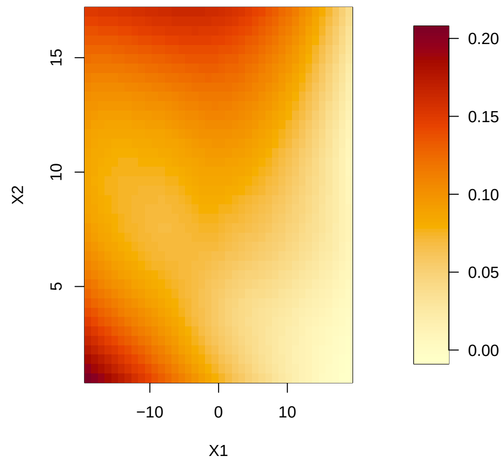


Figure 3: A smoothed heat map of $\bar{Y}$ over the grid of the reception locations of throw-ins. The point $(X1 = 0,\ X2 = 0)$ refers to the throw-in location.

References

[1] Austin, P.C. (2011). *An introduction to propensity score methods for reducing the effects of confounding in observational studies.* Multivariate Behavioral Research, 46, 399-424.

[2] Blank, D. (2012). *Soccer IQ.* www.soccerpoet.com

[3] Epasinghege Dona, N. and Swartz, T.B. (2023a). *A causal investigation of pace of play in soccer.* Statistica Applicata - Italian Journal of Applied Statistics, To appear.

[4] Epasinghege Dona, N. and Swartz, T.B. (2023b). *A causal investigation of throw-ins in soccer.* Manuscript under Review.

[5] Gudmundsson, J. and Horton, M. (2017). *Spatio-temporal analysis of team sports.* ACM Computing Surveys, 50(2), Article 22.

[6] Pearl, J. (2009) *Causality: Models, Reasoning and Inference, Second Edition.* Cambridge University Press: Cambridge.

[7] Wu, Y., Danielson, A., Hu, J. and Swartz, T.B. (2021) *A contextual analysis of crossing the ball in soccer.* Journal of Quantitative Analysis in Sports, **17**, 57-66.

Decision support for sports choices based on performance measurement using the generalized Thurstone method: the model and a case study

László Gyarmati*, László Edvy**, Csaba Mihálykó***, Éva Orbán-Mihálykó****

* Department of Mathematics, University of Pannonia, Hungary; Email: gyarmati.laszlo@phd.mik.uni-pannon.hu
** Physical Education and Sports Institute, University of Pannonia, Hungary; Email: edvy.laszlo@uni-pannon.hu
*** Department of Mathematics, University of Pannonia, Hungary; Email: mihalyko.csaba@mik.uni-pannon.hu
**** Department of Mathematics, University of Pannonia, Hungary; Email: orban.eva@mik.uni-pannon.hu

Abstract

Paired/pairwise comparisons methods are frequently used in sports evaluations. One of them is the Thurstone method with more than two options. In this presentation, we extend the application of the method for decision support in sport selection. We present a model to compute weights for ranked assessment criteria based on coaches' opinions and, combined with the performance of prospective athletes who choose a sport in the assessment, to determine objectively and quantifiably which sport is most likely to be recommended to a given individual. This method allows us to measure the strength of each athlete objectively and compare them across age groups. Extending this method, it can be used not only for sport selection, but also for making various complex decisions when there are several possible outcomes to choose from. As the result of the research, we have generalized the set of criteria and applied it in a specific case study of athletics sport choice.

1 Introduction

Pairwise comparisons are often used in various rankings, decision-making processes, and determining weights for characterizing the importance of the attributes. When making simple decisions, and the number of possible choices is small, there may not be a need to rely on decision support methods, but for more complex ones, the support provided by such methods can be valuable. It can be applied in sports, too [2, 6].

One of the most commonly used methods is the Analytic Hierarchy Process (AHP) [4]. Its main advantages are the possibility of several outcomes for choice and the opportunity of multi-level decision. We wanted preserve these advantages but we wanted to create a method which relies on stochastic aspects.

The Thurstone method stands out among other pairwise comparison methods with stochastic background, and it is highly applicable even for small data sets. It does not require performing every possible comparison and can also provide probabilities for the results of future comparison outcomes [6, 7].

Expanding upon this method, we developed our own approach and applied it as a case study for evaluating the results of athletes. We then compared the obtained results with those provided by coaches and the actual competition results.

2 The model

The model we have created consists of two parts. In the first part, we evaluate the importance of specific attributes (based on sports disciplines) and assign weights to them.

In the second part, we determine the values for the objects to be evaluated (individuals) corresponding to the previously assessed attributes. We assign these values and then multiply them by the corresponding weights associated with each attribute. This process yields the overall evaluation score for the individual.

2.1 First part of the model: assigning weights to attributes

In the Thurstone method allowing more than two options [5], we consider latent random variables, denoted by ξ_i, behind the attributes. We compare the attributes in pairs. Comparing two attributes, the result of the comparison is based on the difference between the latent random variables

$$\xi_i - \xi_j = m_i - m_j + \eta_{i,j}, i = 1, ..., n, j = 1, ..., n, i \neq j, \tag{1}$$

where $E(\xi_i) = m_i$ and $\eta_{i,j}$ is the random error. m_i represents the average strength of the attributes being compared. We applied the Thurstone method, therefore $\eta_{i,j}$ are supposed to be independent identically distributed random variables, with standard Gauss distribution.

We allowed for 7 options (much worse, worse, slightly worse, same, slightly better, better, much better), denoted as C_1, C_2,...,C_7. We assign intervals, I_i, to each of these options based on Figure 1. The points separating the intervals are sequentially denoted as d_1, d_2, d_3.

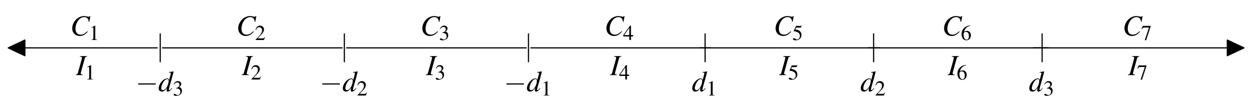


Figure 1: The options and the intervals belonging to them

Let $A_{i,j,k}$ represent the number of decisions C_k, that is the difference $\xi_i - \xi_j$ falling into the k-th interval. Of course, due to the symmetry, $A_{i,j,k} = A_{j,i,8-k}$.

The maximum likelihood estimation of the parameters is the argument at the maximal value of the likelihood function 2 with respect to $\underline{m} = (m_1, ..., m_n)$ and $0 < d_1 < d_2 < d_3$.

$$L(A|m_1, m_2, ..., m_n, d_1, d_2, d_3) = \prod_{k=1}^{7} \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \left(P(\xi_i - \xi_j \in I_k)\right)^{A_{i,j,k}} \tag{2}$$

In case of a 7-option model, the following conditions assure the existence and the uniqueness of the maximizer, supposing $m_1 = 0$:

- The objects compared are represented as vertices of a graph, with a directed edge pointing towards the better option in the case of an extreme (much better) decision, and a bidirectional edge for non-extreme decisions. The resulting graph has to be strongly connected.
- Each decision (much better, better, slightly better, equal, etc.) has to be existed.
- Finally, there is a directed cycle in the graph along much better edges.

During our work, the compared attributes were the physical aspects, and the trainers' opinions provided the results of the comparisons. After evaluating the opinions, we can form weights from the obtained expected values using the following formula:

$$\underline{\widehat{w}} = \left(\frac{\exp(m_1)}{\sum_{i=1}^{n} \exp(m_i)}, ..., \frac{\exp(m_n)}{\sum_{i=1}^{n} \exp(m_i)} \right). \tag{3}$$

2.2 Second part of the model: evaluation of individuals

For each individual, we determine the extent to which he/she fulfills each attribute, separately, by assigning a number between 0 and 1. For example, if a specific attribute has a minimum value of 5 and a maximum value of 10, if the object achieves 5 or below, we assign it a value of 0. If it achieves 10 or above, we assign it the maximum value of 1. Intermediate values are determined along a linear function. For instance, if the value is 7, we evaluate that the individual fulfills the attribute to a degree of 0.4.

The measures obtained for the individuals, which we want to evaluate using the above procedure, are multiplied by the weights previously assigned to the corresponding attributes, and these products are summed up. The resulting score of the individual is thus a number between 0 and 1, where a higher value indicates better fulfillment of the necessary requirements.

3 Application of the model

As a case study, we measured the performance of athletes in different sports disciplines using the method. We aimed to objectively assess the athletes' performance. We developed an application to facilitate the classification of athletes into three sports disciplines: throwing, running, and jumping/sprinting. This application is capable of quantifying the athletes' strengths in each sports discipline.

In Hungary, since 2014, the NETFIT survey [3] has been used to assess the fitness of students. With this 11-item survey, a fairly accurate picture can be obtained regarding the fitness of individual students/athletes, whether he/she involves short or long-term exertion of force, or force exerted by the upper or lower extremities. Similar measures are used in other countries, the first being the FITNESSGRAM® from 1977 [1].

We conducted the NETFIT surveys on young athletes, and based on the results of these surveys, we evaluated the performance of each young athlete with a number ranging from 0 to 1.

The attributes (which were elements of the NETFIT survey) were compared by coaches for each sports discipline. At least three coaches performed the comparisons for each sports discipline, and we calculated the weights for the attributes in each sports discipline based on these comparisons.

The weights for the attributes in each sports discipline are shown in Table 1. The table clearly shows that the attributes most closely associated with each sports discipline dominate. In the throwing discipline, the small ball throw and standing long jump tests are crucial. It may be surprising at first that the standing long jump is important, but in events like hammer throwing and javelin throwing, it is essential to have the ability to generate large forces quickly with the lower extremities, thereby accelerating the sports equipment to be thrown.

Table 1: Weights of Attributes in Surveys Regarding Disciplines

Test name	throwing	running	jumping/sprinting
paced curl-up test	0.063	0.028	0.010
trunklift test	0.062	0.027	0.008
paced push-up test	0.049	0.017	0.009
handgrip strength test	0.061	0.045	0.057
standing long jump test	0.247	0.044	0.232
back-saver sit and reach test	0.038	0.053	0.020
20m pacer test	0.046	0.578	0.020
60m with standing start	0.128	0.051	0.555
small ball throwing	0.213	0.019	0.042
body weight	0.034	0.064	0.009
height	0.059	0.074	0.038
sum of weights	1	1	1

For runners (including middle and long-distance runners), the 20m pacer test is the most important, indicating their capacity for long-term exertion of force. In the jumping/sprinting discipline, explosiveness of the lower extremities is crucial, making the 60m with standing start and standing long jump tests the most important. To determine the importance of tests in each sports discipline more accurately, more coach comparisons are needed, and if possible, further breakdown of the existing sports disciplines could be considered.

4 Results

We verified the reliability of the method in two ways. In the first scenario, we conducted tests on athletes who had not yet been assigned to a specific discipline and asked coaches to indicate which discipline they believed the athletes should be placed in. We then compared their responses with the results obtained from our method.

A total of 18 athletes participated in this test, and they exhibited significant diversity in terms of both age and performance level. It is important to emphasize this, because a 12-year-old may complete a 60-meter race much faster than a 6-year-old. Therefore, we had to create different scales to represent the strength of their results. As a result, we divided the participants into two groups based on their performance (see Table 2).

Out of the 5 older athletes, we correctly identified 4, and out of the 13 younger athletes, we correctly identified 12. In one case, we had to determine the orientation of a 6-year-old, which is particularly challenging and not easily achievable at such a young age. Moreover, his/her scores are very low in every sports,and there is only a slight difference (0.01) between the scores belonging to throwing and running.

In summary, out of the 18 cases, we made recommendations that aligned with the coaches' opinions in 16 cases, resulting in an 89% agreement rate, which can be considered quite good.

The other way of the checking the reliability of the method is the investigation of the relationship between the assigned scores for already categorized athletes and their official rankings. We refer to our own ranking

Table 2: Results of those awaiting classification into a sports discipline

birth year	name	throwing	running	jumping/sprinting	coach recommendation	method recommendation
2011	athlete 1	0.506	0.483	0.637	jumping/sprinting	jumping/sprinting
2012	athlete 2	0.280	0.321	0.427	running	jumping/sprinting
2012	athlete 3	0.409	0.632	0.508	running	running
2012	athlete 4	0.447	0.529	0.605	jumping/sprinting	jumping/sprinting
2012	athlete 5	0.116	0.610	0.047	running	running
2013	athlete 6	0.423	0.592	0.576	running	running
2014	athlete 7	0.558	0.825	0.612	running	running
2014	athlete 8	0.515	0.794	0.568	running	running
2014	athlete 9	0.628	0.317	0.641	jumping/sprinting	jumping/sprinting
2014	athlete 10	0.258	0.232	0.259	jumping/sprinting	jumping/sprinting
2014	athlete 11	0.293	0.221	0.210	throwing	throwing
2015	athlete 12	0.362	0.191	0.492	jumping/sprinting	jumping/sprinting
2015	athlete 13	0.502	0.239	0.591	jumping/sprinting	jumping/sprinting
2015	athlete 14	0.124	0.097	0.220	jumping/sprinting	jumping/sprinting
2015	athlete 15	0.049	0.032	0.146	jumping/sprinting	jumping/sprinting
2016	athlete 16	0.097	0.096	0.032	throwing	throwing
2017	athlete 17	0.068	0.058	0.054	running	throwing
2017	athlete 18	0.018	0.025	0.009	running	running

as the "NETFIT ranking" from now on. In the case of male athletes (see at the top of the Table 3), we had two participants (long jump), and their rankings in both systems matched. Furthermore, we observed that the athlete we rated significantly higher in our system indeed achieved much better results compared to the other athlete in their respective discipline.

Table 3: Results of male and female competitors

name	throwing	running	jumping/sprinting	official ranking	NETFIT ranking
male 1	0.609	0.661	0.819	5.67m	1
male 2	0.671	0.570	0.683	5.05m	2
female 1	0.634	0.495	0.754	13.57s	1
female 2	0.562	0.624	0.666	13.68s	2
female 3	0.452	0.536	0.572	14.33s	3
female 4	0.439	0.432	0.550	13.90s	4

In the case of female athletes (see at the bottom of the Table 3), the two rankings were very similar to each other (100m race). There was only one instance where a swap occurred between the 3rd and 4th positions. In this case, the scores we have given were very close to each other. To be more convincing, we will evaluate any more athletes in the future.

Based on these results we can see that this method is appropriate for predicting the performance of our athletes, making it useful for determining the direction of improvement for each athlete based on the NETFIT results, ultimately leading to better competition results. This is particularly true when dealing with multi-faceted events that require multiple strengths, such as decathlons or other events with diverse skill requirements.

The accuracy of the method for determining the athletes' strengths can be further enhanced by using a function that is more suitable than a linear function within the minimum and maximum values of the scale we applied. This function can be for example the sigmoid function.

5 Conclusions

Throughout our work, we successfully developed a method which is suitable for making complex decisions when multiple factors are involved. We created a method that ranks these factors, assigns weights to them, and evaluates their results in order to assess the "goodness" of complex decisions.

As a case study, we developed a software which is capable of objectively measuring athletes' performance in their respective disciplines. This method proved to be useful not only for determining the appropriate discipline for individual athletes but also for ranking them within their own disciplines without relying solely on competition results. This can greatly assist coaches in objectively assessing the progress of their athletes and identifying the specific areas of development that can contribute to better rankings.

The developed method can be applied to other situations as well, such as assisting in career choices for students or evaluating various tenders, providing decision support in diverse scenarios.

Acknowledgements

The research was supported by the ÚNKP-22-3 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. L. Gyarmati thanks the support.

References

[1] Corbin, C.B.–Lambdin, D.D.–Mahar, M.T.–Roberts, G.–Pangrazi, R.P. (2013) *Why Test? Effective Use of Fitness and Activity Assessments.* In: Plowman, S.–Meredith, M. 2013. Fitnessgram/Activitygram reference guide. Dallas, TX: The Cooper Institute.

[2] Csató, L. (2013) *Ranking by pairwise comparisons for Swiss-system tournaments.* Central European Journal of Operations Research, 21, 783-803.

[3] FitBack (2023) *NETFIT - Hungarian best practice* https://www.fitbackeurope.eu/en-us/monitoring-fitness/best-practice/hungary-netfit Accessed May 23, 2023.

[4] Forman, E.H. (1990) *Multi Criteria Decision Making and the Analytic Hierarchy Process.* In: Bana e Costa, C.A. (eds) Readings in Multiple Criteria Decision Aid. Springer, Berlin, Heidelberg, pp. 295–318.

[5] Orbán-Mihálykó, É., Mihálykó, C., Koltay, L. (2019). *Incomplete paired comparisons in case of multiple choice and general log-concave probability density functions.* Central European Journal of Operations Research, 27(1), pp. 515-532.

[6] Orbán-Mihálykó, É., Mihálykó, Cs., Gyarmati, L., (2022). *Application of the Generalized Thurstone Method for Evaluations of Sports Tournaments' Results.* Knowledge, 2(1), pp. 157-166.

[7] Orbán-Mihálykó, É., Mihálykó, Cs., Gyarmati, L., (2023) *Evaluating the capacity of paired comparison methods to aggregate rankings of separate groups.* Central European Journal of Operations Research, pp. 1–21.

Performance monitoring in anti-doping with Bayesian longitudinal models

Jim Griffin*, Laurentiu C. Hinoveanu**, James Hopker***

* Department of Statistical Science, University College London, London, UK, j.griffin@ucl.ac.uk

** School of Sport and Exercise Sciences, University of Kent, Canterbury, UK, lh460@kent.ac.uk

*** School of Sport and Exercise Sciences, University of Kent, Canterbury, UK, J.G.Hopker@kent.ac.uk

Abstract

In the fight against doping, there is an increasing need to develop methods which allow one to allocate testing resources sensibly. As the primary reason for doping is the improvement of athletic performance, it is reasonable to suggest that monitoring an individual's competition results on a longitudinal basis may reveal suspicious performance improvements. This work is an extension of a recently published performance model which aims to distinguish between normal or expected rates of progression and those caused by doping. We build a Bayesian spline model, which also allows for skewed or heavy-tailed data. These assumptions lead to more robust estimators in the presence of poor performances. We use our model to identify changes in the career performance trajectory of an athlete that may not be consistent with their age-matched cohort. These athletes can be flagged as individuals who may be at greater risk for doping and warrant follow-up investigation. We evaluate the performance of this approach on a data set comprised of male weightlifters.

1 Introduction

The current level of prevalence of doping in elite sports is unknown. Research studies involving anonymous athlete randomised response techniques estimate the prevalence of doping within a 12-month period to be between 20 and 62% across a range of elite sports (de Hon et al. 2015, Ulrich et al. 2018). Despite the number of blood and urine samples taken from athletes across all sports remaining relatively consistent, with approximately 241,430 taken in 2021 (267,645 in 2012, 278,047 in 2019), the percentage of those samples returning adverse analytical findings is falling (1.76% in 2012, 0.82% in 2020, 0.77% in 2021) (WADA 2023). Therefore, questions can be raised about the efficiency of the current anti-doping policy and testing strategies of anti-doping organizations in identifying the right athletes and testing them at the right time. As a consequence, there is a need to gather additional information on athletes to provide a forensic-style intelligence-led approach to anti-doping. As the aim of any doping regime is to improve sporting performance, it has been suggested performance data may be useful in strengthening the sensitivity and applicability of the current methods in the fight against doping. Indeed, Schumacher & Pottgiesser (2009) have previously demonstrated that seasonal world best performances increase with the emergence of new potent doping agents, and reduce when novel anti-doping tests are introduced.

In the literature, there appear to be two main approaches to the statistical analysis of athletic performance over time, namely a time series analysis of performances (Stephenson & Tawn 2013, Gao et al. 2020) or longitudinal models (Stevenson & Brewer 2021, Egidi & Gabry 2018, Hopker et al. 2020). The main benefit of using the longitudinal models is that they account for correlation in the observed performance of each athlete. In this paper, we extend our previous work of performance profiling of athletes (Griffin et al. 2022) using Bayesian longitudinal models within an anti-doping context. The estimates from our model will be used to build performance profiles and, through the introduction of a term called *delta excess performance*, which affords the ability to identify sudden major improvements and/or sustained periods of high performance that might be indicative of athlete doping. In turn, we use the measure of delta excess performance on historical performance data from international weightlifting competitions to demonstrate it can be used to identify athletes with and without previous anti-doping rule violations (ADRVs).

2 Methodology

We have adopted the notation used in Griffin et al. (2022) for consistency. The model utilized in this paper is a Bayesian spline model, and the errors follow a skew t distribution (Azzalini 2013). We have selected this distribution as it allows more flexibility when dealing with real data. Firstly, we can control the departure from the classical symmetric bell-shaped look of normally distributed data through a skewness (α) parameter and through a degree of freedom (ν) parameter, we may allow heavier tails.

The performance of athlete i (out of a population of M athletes) at age j denoted as $y_{i,j}$ is

$$y_{i,j} = g(t_{i,j}) + f_i(t_{i,j}) + x_{i,j}\,\zeta + \varepsilon_{i,j}, \quad i = 1,\ldots,M, \quad j = 1,\ldots,n_i, \tag{1}$$

where:

- the *population performance trajectory* is labeled as $g(t)$ and the individual's *excess performance trajectory* is $f_i(t)$.
- $x_{i,j}$ are population-level covariates that may influence performance like wind speed and direction in an outdoor sprinting event, track altitude, etc.
- ζ are population-level regression coefficients and $\varepsilon_{i,j}$ are observation errors following a skew t distribution.
- $g(t_{i,j}) = \sum_{k=0}^{4} \eta_k (t_{i,j} - \bar{t})^k$ with $\bar{t}$ being the mean age across all athletes and $t_{i,j}$ representing the age of individual i during observation j.
- $f_i(t) = \theta_i + s_{i,j}\boldsymbol{\beta}_i$ with:
  - θ_i is the baseline excess performance.
  - $s_{i,j} = ((t_{i,j} - m_1)^+, \ldots, (t_{i,j} - m_K)^+)$ where $(t_{i,j} - m_1)^+ = \max\{0, t_{i,j} - m_1\}$ and so on.
  - $m_1, \ldots, m_K$ are equally spaced knots between the youngest and oldest among the performing athletes and $K = 100$.
  - $\boldsymbol{\beta}_i$ is a K-dimensional vector of individual-level regression coefficients.

As seen in the works of Frühwirth-Schnatter & Pyne (2010), we may use a Bayesian hierarchical model to represent the skew t distribution, that is:

$$\begin{aligned}
y_{i,j} &= g(t_{i,j}) + f_i(t_{i,j}) + x_{i,j}\,\zeta + \frac{\alpha}{\sqrt{1+\alpha^2}} z_{i,j} + \varepsilon^{\star}_{i,j} \\
\varepsilon^{\star}_{i,j} &\sim \text{Normal}\left(0, \frac{\sigma_i^2}{w_{i,j}(1+\alpha^2)}\right) \\
z_{i,j} &\sim \text{TruncatedNormal}_{[0,\infty)}\left(0, \frac{\sigma_i^2}{w_{i,j}}\right) \\
w_{i,j} &\sim \text{Gamma}(\nu/2, \nu/2)
\end{aligned}$$

The hierarchical representation from above provides a clear path on how to apply a *Markov chain Monte Carlo* (MCMC) scheme to estimate the necessary parameters for our athlete's performance model. More details are available by following the contents of Griffin et al. (2022). Through the respective adaptive MCMC algorithm, we can estimate the performance changes that happen across an individual's career. The aforementioned excess performance captures this. Let us denote with $\Delta_{i,j}^{T} = f_i(t_{i,j}) - f_i(t_{i,j-T})$ the *delta excess performance* at time j and over period T. Across this paper, we take $T = 1$ year. We want to compare an individual's delta excess performance with its population distribution, which is estimated by using the distribution of the posterior medians of delta excess performances at a specific age instance.

Let us denote the 60% percentile at integer age j for the posterior medians as QM_j. The posterior probability score for individual i at age j symbolised by $PPS_{i,j}$ is

$$PPS_{i,j} = \mathbb{P}(\Delta_{i,j}^{1} > QM_j \mid D),$$

where D represents the available data. This $PPS_{i,j}$ can be estimated from the Markov chain Monte Carlo output using

$$\frac{\#\text{ elements of }\Delta_{i,j}^{1} > QM_j}{L},$$

with L being the chain length.

3 Data and Results

3.1 Data

We utilise men's weightlifting performances to outline our method. The source of this data set is the International Weightlifting Federation website. It encompasses 1,609 male lifters with at least five competition results over the period 24th April 1998 to 24th May 2021. This means there are 14,557 observations. The age interval of the considered athletes is from 11 to 43 years old. As a performance, we have considered the **total** lift score, which is made of the **snatch** and **clean and jerk** scores. In Figure 1, we can see the estimated population trajectory for this data set. Moreover, as the data set is comprised of observations in various weight categories, we need a way to put all these performances on the same scale. This is done through the Sinclair totals (`https://iwf.sport/wp-content/uploads/downloads/2017/01/Sinclair_Coefficients_2017.pdf`), which are an athlete's modified total lift score by a specific coefficient called the Sinclair coefficient. This coefficient is characterised by a non-linear relationship between the individual's body weight when he was obtaining the respective lift score and the body weight of the record holder from the heaviest weight class. The Sinclair total is the adjusted lift total that the individual may lift if he is part of the same weight class as the total score record holder.

3.2 Results

In Figure 1, we observe that the peak of the population performance trajectory is obtained around ages 24-25. Furthermore, we see a faster increase in performance improvement amongst ages up to 20 compared to all other ages older than that.

After we let the MCMC algorithm developed by Griffin et al. (2022) run for 10000 iterations, we may use our Bayesian spline estimates from it to compute the delta excess performance measure and *PPS*. In Figure 2, we see the two empirical cdfs (ecdfs) for two groups of athletes in the age interval 24-25 over their yearly delta excess posterior probability scores with the area between the reference curve for athletes without an ADRV (blue line) and the curve for athletes with an ADRV (red line). The difference between both lines illustrates the discriminative power of the model at different posterior probability scores. It is possible to identify that the greatest area between the ecdf curves occurs at a posterior probability score of 0.55, suggesting this threshold could be used to indicate athletes with performances that are at greater risk of doping. Figure 3 illustrates how this probability threshold could subsequently be applied to longitudinal profiles from individual athletes with and without historical ADRVs. This is further seen in the comparison plots between an individual with a doping offence (top row) and one without in. In the subplots from column

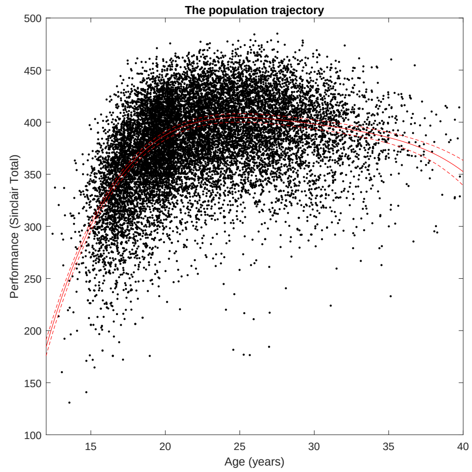


Figure 1: The population performance trajectory for the male weightlifters data set.

IV, our method would highlight the individual with ADRV as suspicious even before he was discovered to have used banned substances.

4 Conclusions

This paper illustrated how a Bayesian longitudinal model could be applied to construct performance profiles for weightlifting athletes. Moreover, how by using the concept of delta excess performance, extreme changes in career trajectories can be identified and used to flag athletes at greater risk of doping.

References

[1] Azzalini, A. (2013), *The Skew-Normal and Related Families*, Institute of Mathematical Statistics Monographs, Cambridge University Press.

[2] de Hon, O., Kuipers, H. & van Bottenburg M. (2015), 'Prevalence of doping use in elite sports: a review of numbers and methods', *Sports Medicine* **45**, 57–69.

[3] Egidi, L. & Gabry, J. (2018), 'Bayesian hierarchical models for predicting individual performance in soccer', *Journal of Quantitative Analysis in Sports* **14**, 143–157.

[4] Frühwirth-Schnatter, S. & Pyne, S. (2010), 'Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-t distributions', *Biostatistics* **11**, 317–336.

[5] Gao, Z., Li, Y. & Wang, Z. (2020), 'Restoring the real world records in Men's swimming without high-tech swimsuits', *Journal of Quantitative Analysis in Sports* **16**, 291–300.

[6] Griffin, J. E., Hinoveanu, L. C. & Hopker, J. G. (2022), 'Bayesian modelling of elite sporting performance with large databases', *Journal of Quantitative Analysis in Sports* **18**(4), 253–268.
URL: *https://doi.org/10.1515/jqas-2021-0112*

[7] Hopker, J., Griffin, J., Brookhouse, J., Peters, J., Schumacher, Y. O. & Iljukov, S. (2020), 'Performance profiling as an intelligence-led approach to antidoping in sports', *Drug Testing and Analysis* **12**(3), 402–409.

[8] Schumacher, Y. O. & Pottgiesser, T. (2009), 'Performance profiling: A role for sport science in the fight against doping?', *International Journal of Sports Physiology and Performance* **4**(1), 129 – 133.

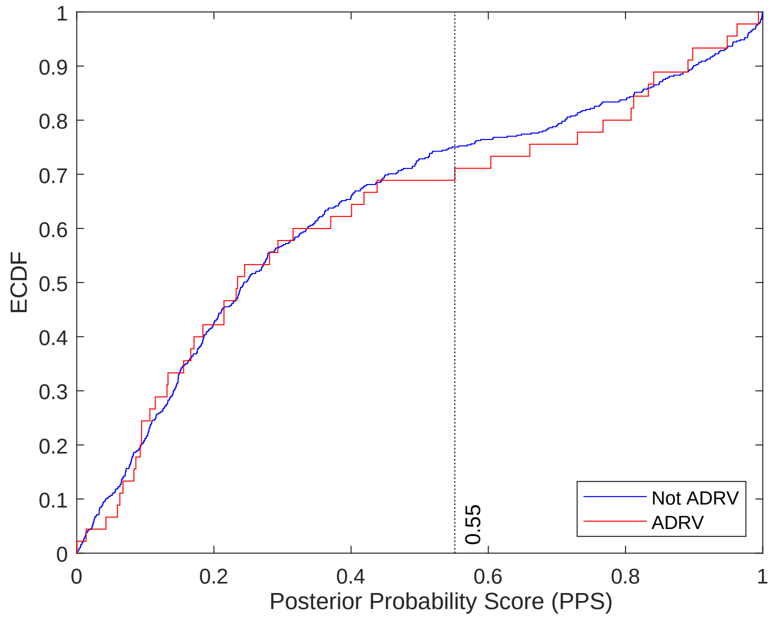


Figure 2: The empirical CDFs for the athletes who have an ADRV (red) and the ones without (blue).

[9] Stephenson, A. G. & Tawn, J. A. (2013), 'Determining the Best Track Performances of All Time Using a Conceptual Population Model for all Athletics Records', *Journal of Quantitative Analysis in Sports* **9**, 67–76.

[10] Stevenson, O. G. & Brewer, B. J. (2021), 'Finding you feet: a Gaussian process model for estimating the abilities of batsmen in test cricket', *Journal of the Royal Statistical Society: Series C* **70**, 481–506.

[11] Ulrich, R., Pope, H. G., Cléret, L., Petróczi, A., Nepusz, T., Schaffer, J., Kanayama, G., Comstock, R. & Simon, P. (2018), 'Doping in two elite athletics competitions assessed by randomized-response surveys.', *Sports Medicine* **48**, 211–219.

[12] WADA (2023), '2021 anti-doping testing figures'. URL: `https://www.wada-ama.org/sites/default/files/2023-01/2021_anti-doping_testing_figures_en.pdf`.

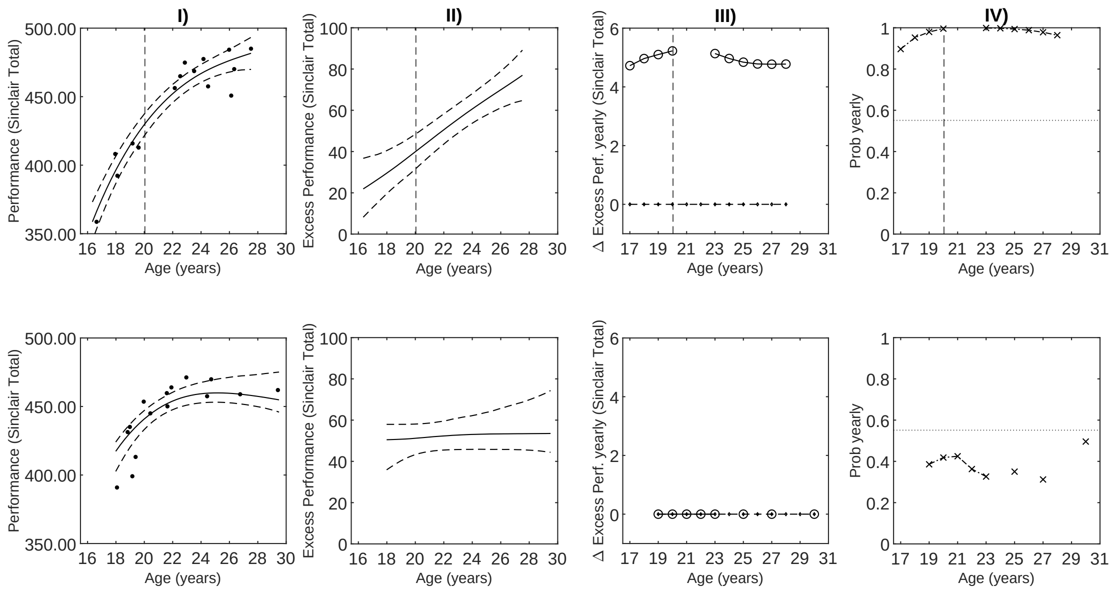


Figure 3: Individual athlete performance trajectories for two male weightlifting athletes, one with an ADRV (top row, ADRV age is shown as a dashed vertical line), and one without an ADRV (bottom row). Column I) Athlete's estimated performance trajectory line (solid line) and 95% credible intervals (dashed lines), II) Athlete's estimated excess performance trajectory line (solid line) and 95% credible intervals (dashed lines), III) Athlete's yearly delta excess performance (solid line and empty circle markers) and the population's 60% threshold of 0 (dash-dotted line), IV) Probability of the athlete's yearly delta excess performance exceeding the 60% threshold at the age instance (*PPS* with cross markers and dash-dotted line). The horizontal dotted line represents the probability threshold of 0.55.

Application of genetic algorithm to solve a special team orienteering problem

Benedek Soós* and Anna Ibolya Pózna**

* Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprém, Hungary
email address: soos2002.benedek@gmail.com
** Department of Electrical Engineering and Information Systems, University of Pannonia, Veszprém, Hungary
email address: pozna.anna@mik.uni-pannon.hu

Abstract

In this paper, a special kind of team orienteering problem is presented which is a mixture of regular and score-type orienteering courses. In this problem, the team has to complete a fixed-order course while visiting a set of elective control points too. The task of the team is to distribute the elective control points among the members in order to minimize the time required to complete the course. In this paper, an effective genetic algorithm-based solution is proposed to solve this problem. Different variants of the genetic operators are implemented and their effect on the algorithm's performance are analysed. The solutions provided by the proposed algorithm are evaluated based on the results of real team orienteering competitions.

1 Introduction

Orienteering is an outdoor sport, where the task of the competitor is to visit several control points located in the terrain, with the help of a map and a compass. There are two main types of orienteering courses: in the classic course the order of the control points is fixed and the goal is to finish the course within the shortest time, while in the score course, the aim of the competitor is to visit as many control points as possible in any order within a given time span.

The Hungarian team orienteering competition is a combination of the classic and the score-type courses. There is a mandatory course with fixed order of control points, and there are several elective control points that can be visited in any order between the mandatory control points. All elective control points must be visited by at least one member of the team, while the mandatory course must be completed by all members. The task of the team is to distribute the elective control points between the members such that the time of the last member to finish is as small as possible.

The above defined Hungarian team orienteering problem (HTOP) is similar to the team orienteering problem (TOP) (Chao, Golden, and Wasil 1996) and the multiple travelling salesman problem (MTSP) (Bektas 2006). Both of these problems are multi-level combinatorial optimization problems, where at one level the cities/control points must be assigned to the different salesmen/team members, then they need to find the shortest path within their given set of cities. These problems are known to be NP-hard, which means that finding the optimal solution is difficult and requires high computational effort. Therefore heuristic and meta-heuristic algorithms (such as genetic algorithms, particle swarm optimization, ant colony optimization, etc.) are often preferred to solve these kinds of problems, providing an approximate solution within a reasonable time (Cheikhrouhou and Khoufi 2021). Motivated by the popularity of genetic algorithms in the field, the aim of this study is to develop a genetic algorithm framework, that can be used to solve the HTOP.

2 Problem statement

The exact rules of the competition are the following: There are 3 members in each team. The mandatory control points have to be visited in a fixed order by all members of the team. All elective control points have to be visited by at least one member of the team. The elective control points can be visited between the first and the last mandatory control points. In order to easily identify the different types of control points of the problem, the following notions are used. The m mandatory control points are denoted by capital letters in alphabetical order, which defines the visiting order, too. The n elective control points are denoted by the unique identifiers of the control points. The start and the finish can be denoted by $S1$ and $F1$, respectively.

3 Genetic algorithm framework

In this section, the details of the developed genetic algorithms are presented. At first, a suitable representation of the solutions that fits the structure of the HTOP was need to be designed. Then different variants of genetic operators (crossover, mutation and selection) were developed in order to achieve an efficient solution method.

3.1 Multi-chromosome representation

In routing problems, the solutions are usually represented by vectors, that contain the order of the control points in the route. In similar problems (TOP and MTSP) single-chromosome, two-chromosome and two-part chromosome representations of the solutions are typically used (Singh and Baghel 2009). However, in the current problem, the fixed-ordered mandatory control points must be treated carefully, therefore a multi-chromosome representation is more suitable, where the paths of the members are represented by different vectors, similarly to (Bederina and Hifi 2017). Formally, the solution X of the HTOP is the set of the three vectors of the three members: $X = \{x_1, x_2, x_3\}$, where x_i denotes the route of the ith member. Note, that the solution should satisfy the constraints in Section 2.

3.2 Fitness function

The fitness function of a solution is the result of the team, which is the time of the last member to finish. In order to compute the value of the fitness function, the length of the routes needs to be known. If the distance between any two control points is given in a matrix D, then the length of the route of the ith member R_i can be computed as follows:

$$R(x_i) = \sum_{j=0}^{dim(x_i)-2} D(x_i(j), x_i(j+1)), \quad i = 1,2,3 \tag{1}$$

where $x_i(j)$ denotes the jth element of the vector x_i. Assuming that all members have equal speeds, the fitness function is determined by the longest route among the members: $f(X) = max(R(x_1), R(x_2), R(x_3))$. The objective of the algorithm is to minimize the fitness function.

3.3 Crossover operators

Two different crossover operators, that are called dominant and chromosome crossover were implemented that use different approaches to create new solutions. The operation of them is introduced in the following.

The developed *dominant crossover* is a 2-parent crossover method, that results in one offspring. This crossover swaps the shortest and longest paths of the two parents. The parents can be paired by the best-best or the best-worst method. The former selects two solutions with the overall best two fitness values and the latter selects the solutions with the best and worst fitness values in the current population. The parent including the shortest path becomes the dominant parent. Then the longest route in the second parent is changed to the shortest route in the dominant parent. After that, all missing and duplicate control points in the remaining two routes are fixed by removing the duplicate control points from the other two routes and randomly inserting the missing control points into the shortest of the two routes.

The second developed crossover operator is the *chromosome crossover*, which is similar to the partially mapped crossover (PMX) (Larranaga et al. 1999). In the chromosome crossover method, the crossover is applied to two members (chromosomes) of the same team. With this approach, the set of elective control points remains unchanged, therefore the resulting offspring has no missing control points. The crossover positions are always the mandatory control points in both vectors because the order of them is fixed in all members.

3.4 Mutation operators

During the mutation, a small change is applied to the current solution in order to refine it. Mutation is a unary operator, that takes one individual and modifies it to get a new solution.

The *team mutation* takes one team and randomly swaps two elective control points between the members. This kind of mutation changes the distribution of the elective control points between the members. The method is very simple and fast and can be easily implemented. The drawback is that the lengths of the sections between mandatory control points do not change.

The second mutation method is called *in-route mutation*, which works on one member of the team. The aim of this mutation is to optimize the individual routes of the members by swapping two randomly selected control points in the route. Some pairs of control points are excluded from the selection, to satisfy the constraints (e.g. two mandatory controls cannot be swapped). This mutation may also change the length of the section between two control points. The disadvantage of the in-route mutation is that it requires the checking of the selected control points before swapping them.

3.5 Selection

The role of the selection operator is to choose the individuals that form the next generation, gradually increasing the quality of the population. In the current genetic algorithm framework, *elitist* selection and *tournament* selection methods were implemented (Shukla, Pandey, and Mehrotra 2015). Elitism ensures that the best solutions always survive, but may reduce the diversity of the population. The tournament selection gives a chance for the weaker solutions to be included in the next generation, resulting in a more diverse population.

4 Analysis of the genetic operators

The performances of the implemented genetic operators are evaluated on a real course from the Open Hungarian Team Championship (2022), with 12 mandatory and 34 elective control points. The algorithm with a population size of 200, maximum iteration number of 1000, crossover probability of 1, mutation probability of 0.005 and with different selection, crossover and mutation settings was run on the test course. With each setting, 20 runs were performed, then the results were averaged out to characterize the performance. The performance of the algorithm was measured by the average improvement in the fitness value.

4.1 Parent pairing of the dominant crossover

In case of the dominant crossover method, the parent selection method was analysed first, to determine which pairing method is better. The results of the best-best and the best-worst pairings can be seen in Table 1. It can be seen that the best-worst pairing gives a significantly better improvement in the fitness value, with 43.95%

compared to 38.88%. Therefore the dominant crossover was used with the best-worst parent selection during the following tests.

Parent selection	Initial fitness (average)	Result fitness (average)	Improvement (average)
best-best	24114.09	14529.54	38.88%
best-worst	24107.13	13091.97	43.95%

Table 1: Results of the genetic algorithm with different parent selection methods

4.2 Optimal parameter setting

In the next step, the 8 possible combinations of the 2 crossover, 2 mutation and 2 selection methods were analysed. In the dominant crossover, the best-worst parent selection method was used, as it was proved to be better based on the previous analysis. In the case of the tournament selection, the tournament size was 3, and the winner of the tournament was added to the new population. The results of the genetic algorithm with the 8 possible parameter settings can be seen in Table 2.

Crossover	Mutation	Selection	Initial fitness (average)	Result fitness (average)	Improvement (average)	Best fitness
dominant	team	tournament	24034.57	15615.09	33.28%	13570.2
dominant	team	elitist	24182.47	13860.79	41.57%	13570.2
dominant	in-route	tournament	23867.28	13155.77	43.34%	11925.2
dominant	in-route	elitist	23854.43	12952.87	44.96%	11659.5
chromosome	team	tournament	24096.00	12673.97	46.39%	11051.5
chromosome	team	elitist	24056.35	11838.95	49.81%	10469.3
chromosome	in-route	tournament	24230.80	12984.33	44.51%	11580.4
chromosome	in-route	elitist	24103.36	12846.05	45.04%	11417.5

Table 2: Results of the genetic algorithm with different parameter settings.

It can be seen, that the greatest improvement in the fitness value (49.81%) was achieved with the combination of chromosome crossover, team mutation and elitist selection. The best solution is also provided by this parameter setting, with a fitness value of 10469.3. The dominant crossover with team mutation and tournament selection seems to be the worst combination of the genetic parameters, with only a 33.28% improvement. It can also be noticed, that the elitist selection is always outperforms the tournament selection, and the chromosome crossover always outperforms the dominant crossover with the same settings. However, such clear conclusion cannot be drawn about the mutation operators. The combination of the in-route mutation with the dominant crossover outperforms the team mutation with the same settings. In contrast, the chromosome crossover should be combined with the team mutation in order to get better results.

5 Evaluation of the results

The result of the 2022 men's race is compared to the solution given by the genetic algorithm. The fitness value of the winning team is 10471.50. The course and the route of the team can be seen in Figure 1, where different colors represent different members. The lengths of the routes are 9870.05, 10471.50 and 8468.83.

The best solution of the genetic algorithm with the optimal settings after 1000 iterations (Table 2.) has a fitness value of 10469.37, with route lengths of 10350.67, 10435.25 and 10469.37. This fitness value is

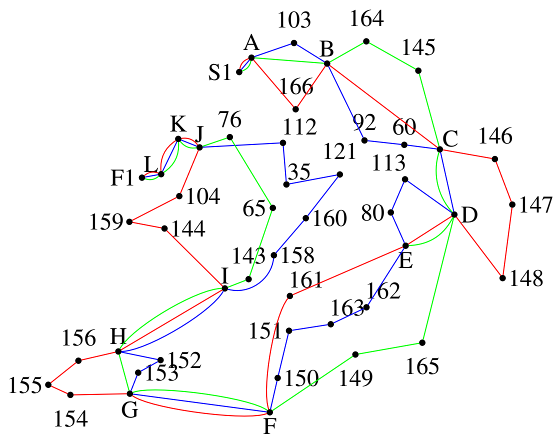


Figure 1: Solution given by the winning team.

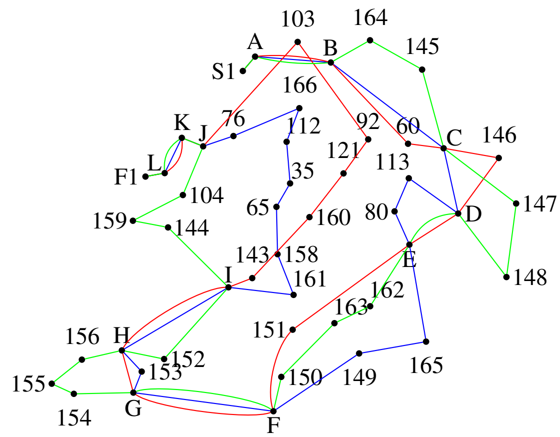


Figure 2: Solution given by the genetic algorithm after 10 000 iterations.

slightly better than the fitness value of the winner. The solution of the genetic algorithm can be further improved by increasing the iteration number. With 10 000 iterations, the fitness value of the solution was 9988.55, with route lengths of 9709.06, 9988.55 and 9955.94, which can be seen in Figure 2. The genetic algorithm provides a more balanced distribution of control points, with only 279 meters difference between the shortest and the longest route, in contrast to the 2002 meters difference in the winner's solution.

Comparing this control distribution with the winner's solution, some similarities and differences can be noticed. Some parts of the distributions are the same in both solutions, e.g. B, 164, 145, C. There are pairs of mandatory control points in both solutions, where no elective control points are visited by any member of the team: F, G and J, K, L. The main difference between the two solutions is the distribution of control points near the start (133, 166 and 92). The winning team visited these control points at the beginning of the course, between mandatory control points A, B and C. In the solution by the genetic algorithm, these control points are visited by one member, at the end of the course. Looking at Figure 2., the red route can be improved by inserting control 103 between A and B, reducing the route length to 9270.49.

6 Conclusions and future work

In this paper, the Hungarian team orienteering problem (a special team orienteering problem) was introduced, which is composed of a fixed mandatory course and numerous elective control points. A meta-heuristic solution method using a genetic algorithm was proposed to find the approximate solution of the given problem. A special multi-chromosome representation of the solutions was introduced and different kinds of crossover and mutation operators were developed to effectively solve the HTOP. Analysis of the operators showed, that the combination of chromosome crossover, team mutation and elitist selection gives the best results.

The developed genetic algorithm was tested on a real course and the results were compared with the solution of the winning team of the race. It can be concluded, that the genetic algorithm can find better and more balanced solutions than the human solution. This phenomenon can be explained by the equal speeds of the members in the algorithm, which rarely occurs in real competitions. In the future, the algorithm can be further improved by including the speeds of the competitors. Besides that, including more information about the terrain that could affect the solution (e.g. vegetation and elevation) can also improve the accuracy of the method.

7 Acknowledgements

Project no. 131501 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the K_19 funding scheme. We thank Zoltán Miháczi for providing us the course data for the analysis.

References

Bederina, H. and M. Hifi (2017). "A hybrid multi-objective evolutionary algorithm for the team orienteering problem". In: *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 0898–0903. DOI: 10.1109/CoDIT.2017.8102710.

Bektas, Tolga (2006). "The multiple traveling salesman problem: an overview of formulations and solution procedures". In: *omega* 34.3, pp. 209–219.

Chao, I-M., B. L. Golden, and E. A. Wasil (1996). "The team orienteering problem". In: *European Journal of Operational Research* 88.3, pp. 464–474.

Cheikhrouhou, Omar and Ines Khoufi (2021). "A comprehensive survey on the Multiple Traveling Salesman Problem: Applications, approaches and taxonomy". In: *Computer Science Review* 40, p. 100369.

Larranaga, Pedro et al. (1999). "Genetic algorithms for the travelling salesman problem: A review of representations and operators". In: *Artificial intelligence review* 13.2, pp. 129–170.

Shukla, Anupriya, Hari Mohan Pandey, and Deepti Mehrotra (2015). "Comparative review of selection techniques in genetic algorithm". In: *2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE)*. IEEE, pp. 515–519.

Singh, A. and A. S. Baghel (2009). "A new grouping genetic algorithm approach to the multiple traveling salesperson problem". In: *Soft Computing* 13, pp. 95–101.

Market characterization of betting in basketball: Evidence from European competitions

Christos Katris[1,2]

[1] Adjunct Lecturer, Department of Mathematics, University of Patras

[2] Customs Officer (Statistician), Independent Authority for Public Revenue, Greece

[1] chriskatris@upatras.gr, [2] c.katris1@aade.gr

Abstract

The scope of this paper is to investigate the nature of betting basketball markets in European competitions. Specifically, we consider the Euroleague 2017-2022 and the Eurocup 2017-2022 from European basketball. The main scope of this paper is to analyze the fairness of the markets. The evidence is based on the Efficient Market Hypothesis (EMH) of Fama and is implemented through a variance-ratio test. Finally, some betting strategies (simple, from Modern Portfolio Theory and from the Risk Management domain) are developing and are considered as an example in order to test whether a player could make money from betting to these markets. The consequences of such a study are whether markets can be characterized according to fairness. Some betting strategies are testing with the use of real data from the European competitions and some conclusions about their effectiveness are extracted.

Keywords: Basketball betting, betting markets, fairness of markets, prediction strategies

1. Introduction

The concept of the efficiency is very important for the analysis of markets. In case there are inefficiencies, in case of correct prediction and measurement, there can be a steady opportunity for the creation of profit. For the financial markets, [1] introduces the renowned Efficient Market Hypothesis (EMH) which, in its weak form, states that markets are efficient in the sense that current prices reflect all the information contained in historical prices. In general, informational efficiency requires that prices represent the best forecasts on the outcome of future events, thus investors cannot achieve a risk-adjusted return in excess of the market by trading on new information. The concept of Market efficiency has application to many kinds of markets, including betting markets. The last decade a number of studies of the bibliography focus on betting markets for the reason that these markets can be seen as a kind of "real world laboratories". The concept of efficiency can be investigated straightforwardly ([2]). A review for both financial and betting markets can be found in [3]. The difference between betting markets and financial markets are firstly the characteristics of the participants, i.e. they are well informed, motivated and experienced and secondly the sports news which offer clean information which is easily shared (difficult the existence of information leakage). Additionally, bets are characterized by a deadline after which their value becomes certain and this fact makes the testing for the market efficiency much simpler.

The basic hypothesis of betting market efficiency is that market prices (i.e. the odds) contain all the relevant historical information and represent the best forecasts on the match outcome probabilities. As a consequence, profits can be generated only by chance and not in a systematic way. Previous papers, e.g. [4], [5], and [6] show that strategies for football matches display abnormal positive returns from betting strategies based on econometric approaches. Additionally, abnormal positive returns have been observed in other sports like American football ([7]; [8]), tennis [9], horserace [10] and Australian Rules football ([11]; [12]). The ability to make abnormal returns implies market inefficiencies. In the field of Basketball there are studies which are mainly related to the NBA and to NCAA. This work is focused on finding the winner (i.e. side line (SL) betting option – spread of points between the favorite and the underdog - is zero). Specifically, the papers [13] and [14] find that SL in the NBA is crucial in predicting the outcome of a game and the market (determination of the point spread between the favorite and the underdog) is influenced by informed traders. In [15] we find the use of the NCAA basketball betting market to find inefficiencies in the determination of differences across various basketball conferences, in [16] is testing whether accuracy in the wagering markets for NCAA Division I men's basketball improves when exist more betting options for a given game. To the best of our knowledge, the literature lacks of contributions which formally test the efficiency of betting markets for online European basketball competitions. However, the concept of efficiency is the same. For testing this effect, variance-ratio tests are employed. Specifically, the Lo-McKinley test [17], the Chow-Denning Multiple Variance Ratio Tests [18], and their bootstrapping [19] and panel [20] versions. Using these tests, we can decide about the efficiency of each market. However, a crucial question is whether the market is mean reverting, persistent or the previous information does not affect this market. This information can be achieved through the Hurst exponent. We use the R/S method to calculate this exponent. The main measure for long memory in time series is the Hurst exponent (H), which was first introduced in the field of hydrology [21] and since then it has been used in a variety of fields. If $0.5 < H < 1$, then we know that the series exhibits long memory and the closer it is to 1, the stronger the long memory. On the contrary, the value $H = 0.5$ indicates absence of long memory, i.e. uncorrelated series or series with autocorrelation function that decays exponentially to zero. If $0 < H < 0.5$ antipersistence is indicated, that is high values are more likely to be followed by low values and vice versa. Antipercistence is the same as mean reverting only if the process has a stable mean [22]. In the experimental work in this paper is considered to be the same. Estimation of the Hurst exponent can be achieved by several methods; however, for our experimental work the R/S method ([17]; [23]) was used. Details for this estimation method can be found in [22]. The rest of the manuscript is as follows: in section 2 is discussed the Efficient Market Hypothesis for the betting markets in basketball, in section 3 are considered the betting strategies and the evaluation of them, in section 4 is presented the experimental work of this paper and section 5 displays the conclusions of this paper.

2. The efficient market hypothesis (EMH) for basketball betting

In this section the scope is to use statistical tests in order to decide if the market is efficient. For this reason, we employ the variance-ratio tests. If the interest is in testing whether or not there is a difference in the group means (this is decided with the equality of group variances), these tests are used. When there are only two groups the test is called a variance ratio test. Generally, the test involves dividing the variance of the one group by the variance of the other group. If this ratio is close to one the conclusion is that the variances of groups cannot be considered to be unequal. The variance ratio (VR) statistic is considered a widely used metric to test for the random-walk hypothesis since the works of [24],[17], [25]. Specifically, in [25] is showed that the VR test is more powerful than either the Dickey-Fuller test or the Box-Pierce Q test for a number of alternatives. Advantages of using the VR test can be found in [26]. If the VR statistic is one, then the series form a random walk, if VR is larger than one then the series display a tendency to form trends and if VR is below one, then the series display a tendency of mean reversion.

3. Betting strategies

In this section, some betting strategies are employed with the scope to display whether a player can make some profits in the European basketball markets. The dimensions of time and the selected league are very important. The considered betting strategies are following:

- ***Wins of the favorites***
- ***Wins of the underdogs***
- ***Wins of Home Teams***
- ***Fibonacci System (predict in favor of underdogs -higher expectation):*** This system utilizes the Fibonacci sequence. The player bets 1 monetary unit at the beginning, the player bet in favor of the underdogs (higher expectation than betting to the favorites) and when wins a bet, the sequence goes down and starts from the beginning. This is the crucial part of the system, i.e. after the winning of a bet how many places should the sequence go down?
- ***The Kelly formula (is considered p=Outsiders Wins/Overall games):*** The betting formula is: $Kf = p - \frac{q}{b}$, where p is the probability of winning and q is the complementary of this probability (q=1-p) and b is the proportion of the bet gained with a win. The player bet in favor of the outsiders and the probability of winning a bet (p) is computed as the wins from the outsiders over the overall games.

The evaluation of betting strategies is performed in terms of ***Profit at the end of the investment, Minimum and Maximum Values of the player overall bankroll*** and ***Variance of a strategy***.

4. Data analysis

The data are coming from the site oddspedia (https://oddspedia.com/) and are the odds for every game in Euroleague and Eurocup. There are considered the highest betting odds for each outcome. There are considered five years of games (from 2017/2018 to 2021/2022) and available odds for them. It is considered a player who only plays one game and bet to all available games. There are 1343 Euroleague and 834 Eurocup games and the bet amount is 100 money units per bet. The (logarithmic) returns are calculated for all the games where a player is betting. The first step is to examine the returns of each market if there are any long-term dynamics. The Hurst Index is calculated for the Euroleague 0.270595 and for the Eurocup 0.3251501. Furthermore, the returns are found to be stationary according to the three versions of ADF test with 0 to 7 lags and display a mean-reverting behavior which makes the hypothesis of a betting streak to be weakening. From Lo-MacKinlay variance Ratio Test the hypothesis of Random Walk of the returns is not supported. We expect that a betting strategy has the ability to offer gains in a regular basis. Next, are considering alternative betting strategies and indeed they offered gains in a player who played with the aforementioned rules (one game per bet and betting to all available games). These betting strategies were evaluated for the 2 competitions and the results are shown in the next Table.

EVALUATING BETTING STRATEGIES (Euroleague 2017-2022)				
BETTING STRATEGY	***MONEY FOR THE INVESTOR***			
	Overall amount	***Minimum amount***	***Maximum amount***	***Variance***
Wins of the favorites	85.539,2	143	85.694,2	13.348,41
Wins of the underdogs	42.696	-535	42.796	41.695,98
Wins of Home Teams	86.820,2	-100	86.820,2	20.627,54
Fibonacci system	50.597,84	-69.310,29	50.598,84	10.570.640
Kelly Criterion (bet to outsiders)	6.046,385	-75,76391	6.060,547	836,2024
EVALUATING BETTING STRATEGIES (Eurocup 2017-2022)				
BETTING STRATEGY	***MONEY FOR THE INVESTOR***			
	Overall amount	***Minimum amount***	***Maximum amount***	***Variance***
Wins of the favorites	55.880	180	55.880	12.744,61
Wins of the underdogs	15.800	-420	16.200	39.006,29
Wins of Home Teams	43.508	34	43652	16.951,99
Fibonacci system	5811,44	-860,86	5.829,44	12.477,3
Kelly Criterion (bet to outsiders)	1723.916	-45,82563	1.767,56	464,3579

5. Summary and conclusions

In this work the scope is to characterize the betting markets of Euroleague and Eurocup basketball competitions as efficient or not. To achieve this, the Lo-MacKinlay variance Ratio Test is employed. Furthermore, to study the dynamics of the markets, the Hurst exponent is calculated through the rescaled range (R/S) analysis. Moreover, we should not be detecting betting strategies which produce constant gains and for this reason we employed different betting strategies. The markets are characterized as stationary and mean reverting and the considered betting strategies found to produce profits for players.

References

[1] Fama, E. F. (1970) *Efficient market hypothesis: A review of theory and empirical work*. Journal of Finance, **25(2)**, 28-30.

[2] Thaler, R. H., & Ziemba, W. T. (1988). *Anomalies: Parimutuel betting markets: Racetracks and lotteries*. Journal of Economic perspectives, **2(2)**, 161-174.

[3] Williams, L. V. (Ed.). (2005). *Information efficiency in financial and betting markets.* Cambridge University Press.
[4] Koopman, S. J., & Lit, R. (2015). *A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League.* Journal of the Royal Statistical Society: Series A (Statistics in Society), **178(1)**, 167-186.
[5] Boshnakov, G., Kharrat, T., & McHale, I. G. (2017). *A bivariate Weibull count model for forecasting association football scores*. International Journal of Forecasting, **33(2)**, 458-466.
[6] Angelini, G., & De Angelis, L. (2017). *PARX model for football match predictions*. Journal of Forecasting, **36(7)**, 795-807.
[7] Glickman, M. E., & Stern, H. S. (2005). *A state-space model for national football league scores. In Anthology of statistics in sports* (pp. 23-33). Society for Industrial and Applied Mathematics.
[8] Boulier, B. L., & Stekler, H. O. (2003). *Predicting the outcomes of National Football League games. International Journal of forecasting*, **19(2)**, 257-270.
[9] McHale, I., & Morton, A. (2011). *A Bradley-Terry type model for forecasting tennis match results*. International Journal of Forecasting, **27(2)**, 619-630.
[10] Lessman, S., Sung, M. C., & Johnson, J. E. (2007). *Adapting least-square support vector regression models to forecast the outcome of horseraces*. The Journal of Prediction Markets, **1(3)**, 169-187.
[11] Grant, A., & Johnstone, D. (2010). *Finding profitable forecast combinations using probability scoring rules*. International Journal of Forecasting, **26(3)**, 498-510.
[12] Ryall, R., & Bedford, A. (2010). *An optimized ratings-based model for forecasting Australian Rules football*. International Journal of Forecasting, **26(3)**, 511-517.
[13] Brown, W. O., & Sauer, R. D. (1993). *Does the basketball market believe in the hot hand? Comment.* The American Economic Review, **83(5)**, 1377-1386.
[14] Gandar, J. M., Dare, W. H., Brown, C. R., & Zuber, R. A. (1998). *Informed traders and price variations in the betting market for professional basketball games*. The Journal of Finance, **53(1)**, 385-401.
[15] Colquitt, L. L., Godwin, N. H., & Caudill, S. B. (2001). *Testing efficiency across markets: Evidence from the NCAA basketball betting market.* Journal of Business Finance & Accounting, **28(1-2)**, 231-248.
[16] Berkowitz, J. P., Depken II, C. A., & Gandar, J. M. (2015). *Information and accuracy in pricing: Evidence from the NCAA men' s basketball betting market.* Journal of Financial Markets, **25**, 16-32.
[17] Lo, A. W., & MacKinlay, A. C. (1988). *Stock market prices do not follow random walks: Evidence from a simple specification test*. The review of financial studies, **1(1)**, 41-66.
[18] Chow, K. V., & Denning, K. C. (1993). *A simple multiple variance ratio test*. Journal of Econometrics, **58(3)**, 385-401.
[19] Kim, J. H. (2006). *Wild bootstrapping variance ratio tests*. Economics letters, **92(1)**, 38-43.
[20] Kim, J. H., & Shamsuddin, A. (2015). *A closer look at return predictability of the US stock market: evidence from new panel variance ratio tests*. Quantitative Finance, **15(9)**, 1501-1514.
[21] Hurst, H. E. (1951). *Long-term storage capacity of reservoirs. Transactions of the American society of civil engineers*, **116(1)**, 770-799.
[22] Peters, E. E. (1994). *Fractal market analysis: applying chaos theory to investment and economics (Vol. 24).* John Wiley & Sons.
[23] Mandelbrot, B. (1972). *Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis.* In Annals of Economic and Social Measurement, Volume 1, Number 3 (pp. 259-290). NBER.
[24] Cochrane, J. H. (1988). *How big is the random walk in GNP?*. *Journal of political economy*, **96(5)**, 893-920.
[25] Lo, A. W., & MacKinlay, A. C. (1989). *The size and power of the variance ratio test in finite samples: A Monte Carlo investigation. Journal of econometrics*, **40(2)**, 203-238.
[26] Cecchetti, S. G., & Lam, P. S. (1994). *Variance-ratio tests: small-sample properties with an application to international output data. Journal of Business & Economic Statistics*, **12(2)**, 177-186.

A stochastic analysis of the 4 x 100 m relay

Niklas Karlsson* and Anders Lunander**

* Örebro University, SE 701 82 Örebro, Sweden; email address: niklas.karlsson@oru.se

** Örebro University, SE 701 82 Örebro, Sweden; email address: anders.lunander@oru.se

Abstract

The baton exchanges are undoubtedly the most critical parts of the 4 × 100 meters relay race. Timing of the outgoing runner is critical. In our paper we analyze the race as a minimization problem under uncertainty. We formulate a stochastic model in which an outgoing runner at the baton exchange cannot perfectly assess the incoming runner's exact location relatively a check mark position, and therefore potentially misjudges the right moment to start running. Also, the team members' daily shape is subject to uncertainty. To understand the effect of these two random variables - incoming runners' distance to checkmark and the daily shape of the running team – we conduct a simulation study to investigate the trade-off between the team's expected race time and their probability of being disqualified due to overrunning the takeover zone. The running profile of the relay team is taken from previous studies in the field. Conditioning on a low disqualification probability, the difference in expected race time is shown to be substantial between teams with different variation in distance assessment and forecasting running performance, respectively.

1 Introduction

The perhaps most spectacular and exiting event in athletics is the 4×100 m relay race. The race captures not only athletes' acceleration capacities and speed endurance but also their ability to work as teams at the baton exchanges. No matter how fast each member of a team is as an individual sprinter, there is always a risk that the team will achieve a bad relay time or even be disqualified due to false baton exchanges. Several factors inherent in the 4×100 m relay are recognized to affect the team's relay time, e.g., choice of lane, starting positions, checkmark positions, takeover positions. Ward-Smith and Radford (2002) and Radford and Ward-Smith (2003) provide a solid mathematical analysis of the relative importance upon running time of varying some factors affecting the relay time. However, there are no stochastic elements in their analysis. The predicted optimal positions of checkmarks rest on the assumption that the outgoing athlete correctly judges when it is time to start accelerating towards the takeover point. Hence, their model is deterministic in the sense that a team does not run the risk of being disqualified when sticking to the predicted positions. Clearly, and admitted by the authors, the use of checkmarks is a hazardous business. A too early departure of the outgoing runner, caused by a misjudgment of the position of the incoming runner, relative to the checkmark, will severely increase the risk of disqualification. Also, a delay of the departure of the outgoing runner likely ensures that the baton exchange takes place inside the takeover zone, but at the cost of a higher total running time. In this study we introduce uncertainty when modeling the optimal positions in the 4×100 m relay. Our primary objective is to investigate the trade-off between the relay team's goal to minimize its total running time and its desire to reduce the risk of failure, due to the outgoing runner's difficulty to correctly judge the right moment to start running. In addition, we also think of the athletes' daily form being stochastic, that is, the athletic performance of the team's members in the race cannot be perfectly predicted by the team's coach. Our numerical analysis is based on the performance profile of an athlete running 100 meters, derived in Ward-Smith (2000). We fit their data into a third-degree polynomial using ordinary least squares (OLS) to obtain a continuous differentiable function reflecting the relation between running time and distance. To understand the effect of the two stochastic components - incoming runners' distance to checkmark and the daily shape of the running team – we conduct a simulation study to investigate the trade-off between the team's expected race time and its probability of being disqualified due to overrunning the takeover zone. However, to facilitate our analysis, we model the 4×100 m relay as a race on a straight track, that is, we exclude the effects of track curvature on running performance.

2 A Deterministic Model for the 4 x 100 m Relay

For a $4 \times 100\ m$ relay race, we define the running time in seconds for the athlete on the ith leg as a function of a covered distance of d meters, $t_i = f_i(d)$.The function is increasing everywhere. It is concave for the acceleration phase where the outgoing runner receives the baton, while it is convex for the range of deceleration where the incoming runner hands over the baton. We define s_i as the starting position of the ith runner, measured in meters from the starting point of the race. Clearly, $s_1 = 0$ is fixed, while s_2, s_3 and s_4 are to be chosen by the relay team within certain intervals, the so-called takeover zones. These zones are represented by the intervals $[80, 110], [180,210]$ and $[280,310]$, respectively. Furthermore, for the ith runner, we define e_i as the finishing position, again measured in meters from the starting point of the race. Here, $e_4 = 400$ is fixed. However, e_1, e_2 and e_3, the positions of the three baton exchanges are, just like s_2, s_3 and s_4, to be chosen by the team within the above intervals. The choice of the values of these six variables determines the race time, $t^* = h(e_1, e_2, e_3, s_2, s_3, s_4)$. The team is looking for the set of values $(e_1, e_2, e_3, s_2, s_3, s_4)$ minimizing the race time. We formulate the optimization problem as

$$\text{minimize}\, h(e_1, e_2, e_3, s_2, s_3, s_4)$$
$$= f_1(e_1)$$
$$+f_2(e_2 - s_2) - f_2(e_1 - s_2)$$
$$+f_3(e_3 - s_3) - f_3(e_2 - s_3)$$
$$+f_4(400 - s_4) - f_4(e_3 - s_4)$$
$$\text{subject to}$$
$$80 \leq s_2 \leq e_{1,} \leq 110,\ 180 \leq s_3 \leq e_{2,} \leq 210,\ 280 \leq\ s_4 \leq e_{3,} \leq 310.$$

The race time is expressed as the sum of the four runners' individual race times carrying the baton along the track. Suppose there is a point $(e_1^*, e_2^*, e_3^*, s_1^*, s_2^*, s_3^*)$ at which h is stationary and for which the inequalities are fulfilled. Then, a necessary and sufficient condition for a minimum is given by the following first order conditions:

$$\frac{\partial h}{\partial e_1} = f_1'(e_1) - f_2'(e_1 - s_2) = 0, \tag{1}$$
$$\frac{\partial h}{\partial e_2} = f_2'(e_2 - s_2) - f_3'(e_2 - s_3) = 0, \tag{2}$$
$$\frac{\partial h}{\partial e_3} = f_3'(e_3 - s_3) - f_4'(e_3 - s_4) = 0, \tag{3}$$
$$\frac{\partial h}{\partial s_2} = -f_2'(e_2 - s_2) + f_2'(e_1 - s_2) = 0, \tag{4}$$
$$\frac{\partial h}{\partial s_3} = -f_3'(e_3 - s_3) + f_3'(e_2 - s_3) = 0, \tag{5}$$
$$\frac{\partial h}{\partial s_4} = -f_4'(400 - s_4) + f_4'(e_3 - s_4) = 0. \tag{6}$$

Besides the variables e_i and s_i, two more variables are of interest for our analysis, the checkmark position and the checkmark distance. The checkmark position for the ith runner, denoted cm_i, $i = 2, 3, 4$, is defined as the position of the incoming runner, measured from the start of the race, at the point of time when the ith outgoing runner is starting to run. Once starting position s_i and exchange position e_{i-1} are determined, cm_i can be solved for by using the equation

$$f_{i-1}(e_{i-1} - s_{i-1}) - f_{i-1}(cm_i - s_{i-1}) = f_i(e_{i-1} - s_i), i = 2, 3, 4. \quad (7)$$

The equation says that the time it takes for the incoming runner to cover the distance from the checkmark position to the exchange position is equal to the time it takes for the outgoing runner to cover the distance from his start position to the exchange position. Finally, the checkmark distance cmd_i is defined as

$$cmd_i = s_i - cm_i, \ \ i = 2, 3, 4. \quad (8)$$

2.2 Empirical Implementation

Our point of departure when empirically determining the optimal start and exchange positions, is the elite male sprint running profile presented in Ward-Smith and Radford (2002), where each of the four runners in the team is supposed to have this profile. Their data is reproduced in columns 2-4 in Table 1.

Table:1 Representative running profile

	Based on Ward-Smith and Radford			Polynomial fitting	
Distance (m) (1)	Time (s) (2)	Time (s) (3)	Speed ($m \cdot s^{-1}$) (4)	Time (s) (5)	Speed ($m \cdot s^{-1}$) (6)
0	0	0.18	0		
10	1.72	1.90	8.85	1.90	9.04
20	2.76	2.94	10.28	2.94	10.19
30	3.70	3.88	10.97	3.88	11.03
40	4.59	4.77	11.35	4.77	11.34
50	5.46	5.64	11.53		
60	6.33	6.51	11.60	6.51	11.67
70	7.19	7.37	11.57	7.37	11.59
80	8.06	8.24	11.49	8.24	11.48
90	8.93	9.11	11.35	9.11	11.35
100	9.82	10.00	11.19	10.00	11.19
110	10.72	10.90	11.00	10.90	11.01
120	11.64	11.82	10.81	11.82	10.81
130	12.57	12.75	10.60	12.75	10.60

Our method to determine the optimal starting and finishing positions, requires a continuous and differentiable function for the team members' running profiles relating the time to cover a certain distance. Therefore, a third-degree polynomial was fitted by OLS to the data. To obtain as good fit as possible, we decided to estimate one function for the acceleration part (0-60 meters) and another function for the deceleration part (60-130 meters). The estimated polynomial for the acceleration part, based on four observations, is given by

$$time = 0.530 + 0.128 \times d - 0.00100 \times d^2 + 0.00000833 \times d^3 \quad (9)$$

and the corresponding estimates for the deceleration part, based on eight observations, is given by

$$time = 1.20 + 0.0865 \times d - 0.0000389 \times d^2 + 0.000000354229 \times d^3. \quad (10)$$

The predicted time is reported in the last two columns in Table 1. Our predictions are overall very good, slightly overestimating the speed at 10 meters. For the minimization problem set out in the previous section, equation (9) is used to model $f_i(\cdot)$ evaluated at $e_{i-1} - s_i$, $i = 2, 3,4$, while equation (10) is used to model $f_1(\cdot)$ at e_1, $f_i(\cdot)$ evaluated at $e_i - s_i$, $i = 2, 3$ and $f_4(\cdot)$ at $400 - s_4$. In addition, a reaction time of 0.18 is added to the first runner. Table 2 shows optimal start and exchange positions as a solution of the minimization problem, both with and without the restrictions implied by the take-over zones.

Table 2. Optimal start and exchange positions with resulting race time, with and without restrictions

Zone restrictions	s_2	e_1	s_3	e_2	s_4	e_3	Race time
Yes	84.1	110	182.1	207.9	280	305.9	36.56
No	93.2	120.0	186.4	213.2	279.7	306.6	36.54

The first and third takeover zones constitute binding restrictions. The first exchange will take place at 110 meters, the upper limit of the first zone, while the fourth runner will start at 280 meters, the lower limit of the third zone. The restrictions affect start and exchange positions to a certain extent, while the effect on race time is marginal, 0.02 seconds. It is also implicitly shown that the distance acceleration phase for the outgoing runners is the same. This holds for both cases, slightly less than 26 meters with restrictions and almost 27 meters without restrictions. These results suggest the speed at the exchange of the outgoing runners be the same for all runners. Table 3 shows the speed at each exchange for the incoming and outgoing runner, as well as the speed at 400 meters for the fourth runner. The first order conditions set out in the previous section is fulfilled for the case with no restrictions. However, with restrictions considered, the speed of the incoming runner at the first exchange is higher than the speed of the outgoing runner, violating condition 1. Also, condition six is also violated, since the fourth runner's speed is higher at the finishing line than at the third exchange.

Table 3. Estimated speed at the exchanges

	Restrictions			
	Yes		No	
Position	Incoming	Outgoing	Incoming	Outgoing
e_1	11.01	10.73	10.81	10.81
e_2	10.73	10.73	10.81	10.81
e_3	10.73	10.73	10.81	10.81
400	10.81		10.81	

3 A Stochastic Model for the 4 x 100 m Relay

First, focusing on the running profile for a certain relay race, we model the time it takes for runner i to run d meters as $T_i = W_i f_i(d)$, where W_i, $i = 1,2,3,4$, is a random variable, here assumed to be $N(\mu_{W_i}, \sigma_{W_i})$, while $f_i(d)$ is the running profile suggested by the coach before the race, being the basis for determining optimal start and exchange positions. The special case where $\mu_{W_i} = 1$ and $\sigma_{W_i} = 0$ represents the coach having perfect information about the runner's daily shape. The case $\mu_{W_i} = 1$ and $\sigma_{W_i} > 0$ means the coach is unbiased, still he sometimes overestimates the runner's performance and sometimes underestimate it. If $\mu_{W_i} > 1$ the coach is systematically overestimating the performance, while if $\mu_{W_i} < 1$ the performance is systematically underestimated. Second, as for possible misjudgment of distance, we define the random variable DS_i as the position of the incoming runner, measured from the start of the race, when the ith outgoing runner starts to run. The difference $Y_i = DS_i - cm_i, i = 2, 3, 4,$ is assumed to follow a $N(\mu_{Y_i}, \sigma_{Y_i})$. The special case where $\mu_{Y_i} = 0$ and $\sigma_{Y_i} = 0$ means the outgoing runner always starts according to plan as the incoming runner passes the checkmark. The case $\mu_{Y_i} = 0$ and $\sigma_{Y_i} > 0$ represents an unbiased judgement of distance, although he sometimes goes too early ($y_i < 0$) and sometimes too late ($y_i > 0$). If $\mu_{Y_i} > 0$ the runner is systematically going too late, while if $\mu_{Y_i} < 0$ he systematically goes too early. As explained above, allowing for these two stochastic elements implies the team's loss is either a worsen race time or a disqualification. The loss function is certainly asymmetric, suggesting the checkmark distance to be shortened to lower the probability of disqualification by overrunning the takeover zone at the expense of worse expected race time. We assume the coach's forecast of the running profile $f_i(d)$, $i = 1, 2, 3, 4$, is given by equations (9) and (10) for all runners where as before, a reaction time of 0.18 is added to the first runner. This forecast forms the basis for the starting positions given in the first row of Table 2, to be denoted s_i^*, $i = 2, 3, 4$. In addition, unbiased forecasting of the running profile is assumed, i.e., the parameter μ_{W_i} will be set equal to 1 for all i. Although a biased distance assessment is likely, we restrict the analysis to cases with no bias assuming $\mu_{Y_i} = 0$ for all $i = 2, 3, 4$. This is motivated by the notion that, in case of bias, the checkmark distance can be corrected accordingly to the presumably known bias $\mu_{Y_i} \neq 0$. To study the effect of the variation of the forecast error, the parameter σ_{W_i} takes on two different values, 0.0025 and 0.005, reflecting various degrees of ability to assess running performance. The variation in the distance assessment measured in meters is varied according to $\sigma_{Y_i} = 0.1(0.2)0.5$. Combined, we end up with $3 \cdot 2 = 6$ cases. For each case, the values of σ_{W_i} and σ_{Y_i}, respectively, are the same for all i. Each case is divided into several cells, where each cell corresponds to a certain set of checkmark distances for the four runners. Define c_i, measured in meters, to be the shortening of the checkmark distance.

For $\sigma_{Y_i} = 0.1(0.2)0.5$, the shortening of the checkmark distance, c_i, $i = 2, 3, 4$, is varied for each of the last three runners according to $0(0.1)1.1$, $0(0.1)1.3$ and $0(0.1)1.6$, respectively, yielding a total of $12^3 = 1728$, $14^3 = 2744$ and $17^3 = 4913$ different sets.

For each cell we conduct 10000 replicates, where each replicate consists of seven independent draws from each of the seven random variables $W_i, i = 1, 2, 3, 4$, and $Y_i, i = 2, 3, 4$. Thus, each replicate represents a simulation of a relay race, delivering one set of running profiles (t_1, t_2, t_3, t_4) and one set of positions of the incoming runner (ds_2, ds_3, ds_4) when the outgoing runner starts to run. The positions are measured from the start of the race and calculated as $ds_i = cm_i + y_i$. These two sets make it possible to determine whether a particular simulated race results in a disqualification or not, and in the latter case, to calculate a race time. This is accomplished by the following sequential scheme:

1) The starting time of the second runner $t_2^{''}$ is determined given t_1 and ds_2.
2) The profiles t_1 and t_2 together with $t_2^{''}$ determines whether an exchange will occur or not. In the latter case the team is disqualified, and we move to next race. In the former case the first exchange position $e_1^{''}$ is calculated and the time for the first runner to carry the baton from start to $e_1^{''}$ is calculated as $t_1^{**} = w_1 f_1(e_1^{''})$.
3) The starting time of the third runner $t_3^{''}$ is determined given $t_2^{''}$, t_2 and ds_3.
4) The profiles t_2 and t_3 together with $t_3^{''}$ determines whether an exchange will occur or not. In the latter case the team is disqualified, and we move to next race. In the former case the second exchange position $e_2^{''}$ is calculated and the time for the second runner to carry the baton from $e_1^{''}$ to $e_2^{''}$ is calculated as $t_2^{**} = w_2 f_2(e_2^{''} - s_2^*) - w_2 f_2(e_1^{''} - s_2^*)$.
5) Steps 3 and 4 are repeated for the third exchange. Here, $t_3^{**} = w_3 f_3(e_3^{''} - s_3^*) - w_3 f_3(e_2^{''} - s_3^*)$.
6) The time for the fourth runner to carry the baton from $e_3^{''}$ to 400 is calculated as $t_4^{**} = w_4 f_4(400 - s_4^*)$.
7) Finally, the team's race time is calculated as $t^{**} = \sum_{i=1}^{4} t_i^{**}$.

Consider Figure 1. Each point of the total of 2744 points represents the probability of disqualification and the expected race time based on 10000 simulations for a certain set (c_2, c_3, c_4), for the case $\sigma_{W_i} = 0.005$ and $\sigma_{Y_i} = 0.3$. Points minimizing the expected race time given a certain disqualification probability are of particular interest. They are tracing out an efficient frontier, showing the best possible trade-off between the expected race time and the probability of disqualification. The price of lowering the disqualification probability is a worse expected race time.

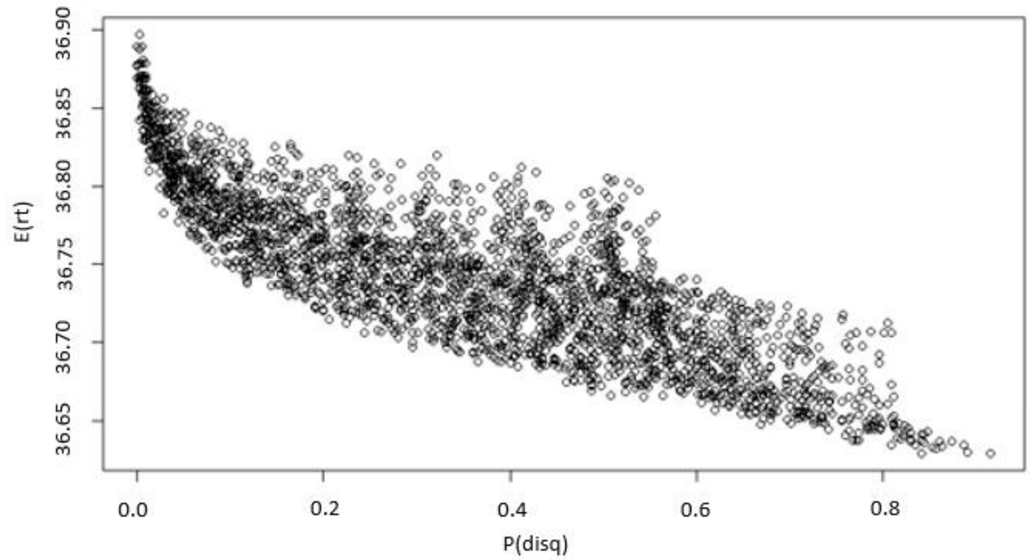


Figure 1. Probability of disqualification and expected race time

4 Summary

Ward-Smith and Radford (2002) and Radford and Ward-Smith (2003) discuss the importance of taking speed reduction on bends into account when deriving optimal start- and exchange positions, as well as check mark distances. Since we assume four straight legs, our results cannot fully be generalized to the real track and field event of $4 \times 100\ m$ relay. By comparing results, start and exchange positions are only marginally affected, while our check mark distances differ percentage wise more and should therefore not be used for normative purposes. Despite this shortcoming of our approach, it is motivated for two reasons. First, it offers an increased understanding of optimizing exchanges in terms of the derived first order conditions about speed. Second, our approach offers an easy way to highlight the importance of distance assessment ability and forecasting running performance skills on the tradeoff between expected race time and probability of disqualification.

References

[1] Radford, P. F. & Ward-Smith, A. J. (2003). The baton exchange during the 4× 100 m relay: a mathematical analysis. *Journal of Sports Sciences*, *21*(6), 493-501.

[2] Ward-Smith, A. J. & Radford, P. F. (2000). Investigation of the kinetics of anaerobic metabolism by analysis of the performance of elite sprinters. *Journal of Biomechanics*, *33*(8), 997-1004.

[3] Ward-Smith, A. J. & Radford, P. F. (2002). A mathematical analysis of the 4× 100 m relay. *Journal of Sports Sciences*, *20*(5), 369-381.

Profitable sports forecasting – Issues in evaluating predictive power and implications for model selection

Fabian Wunderlich*, Marc Garnica Caparrós*, Daniel Memmert*

* Institute of Exercise Training and Sport Informatics, German Sport University Cologne,
Contact: f.wunderlich@dshs-koeln.de

Abstract

Sports forecasting is a diverse field of research and forecasting models are designed for a variety of objectives, including accuracy and profitability as two of the most important. Investigating the profitability of such models in the betting market, however, is subject to specific pitfalls. Based on theoretical considerations as well as large-scale simulation models, we showcase three of these issues: First, profitability and accuracy are clearly different concepts and we argue that profitable models do not necessarily require an excellent model accuracy. Second, we illustrate how the use of profitability measures reduces sample sizes. Consequently, the inherent randomness in the betting returns should be discussed, acknowledged and quantified. Third, we showcase that profitability is not yet sufficiently considered in model specification. Our results suggest that models fitted for profitability are more successful in exploiting bookmaker biases than models fitted for accuracy. This work aims to stimulate further discussion on optimal methods of fitting forecasting models for profitability.

1 Introduction

Researchers have been interested in sports-related forecasts for decades [1, 2]. Given the many different perspectives on the topic, sports forecasting models are investigated for a multitude of reasons. In the literature, models are commonly related to one or several of the following objectives: Testing theories like market efficiency [3] or crowd wisdom [4], understanding human forecasting heuristics [5], using rich datasets from sports to make statistical or machine learning contributions [6], using a predictive model to better understand the mechanisms of the underlying sport [7], developing models capable of generating financial profits at the betting market [8] or meeting the needs of sports broadcasters [9].

In light of this multitude of perspectives, it is not clear from the outset what a “good“ sports forecasting model is. Sports broadcasters might prefer an intuitively understandable model over a complex model; a model capable of handling complex data might have value to the machine learning community, even if not being more profitable than a model based on simpler data; sports bettors naturally should prefer the most profitable over the most accurate model and sports scientists would be interested in the model most accurately representing the mechanisms of the underlying sport. Development and validation of models thus become dependent on the exact objective of the model, which, however, does not seem to be sufficiently addressed in the literature. The present contribution focuses on two of the most important objectives in forecasting, namely accuracy and profitability. While it is state of the art to report measures of accuracy (such as likelihood, Brier scores or rank probability scores) concurrently with measures of profitability (i.e. betting returns of various betting strategies) for probabilistic forecasting models, systematic discussion of the differences between both concepts is rare so far [10, 11].

2 Issues of profitability measures in sports forecasting

2.1 Profitability does not require accuracy

Accurate and profitable forecasting in sports are two different tasks, as the former intends to estimate probabilities close to the true probabilities and the latter intends to exploit bookmaker errors. As such, the intuitive notion that profitable models necessarily require high model accuracy is wrong. We can demonstrate this by a very simple, but intuitive example. Let's assume a tennis match between players A and B with equal odds of 1.9 on each side, where the bookmaker failed to capture the fact that (based on the true probabilities) player A is a 60% favorite to win the match. We further assume a forecaster (i.e. bettor) that estimates the probability of each player winning and bets on a player if, based on its estimation, the expected net win of the bet is positive. Table 1 provides example calculations and Figure 1 illustrates the profitability (measured as net win) and accuracy (measured as rank probability score, RPS) of the forecasters model if varying its estimation for the probability of player A. Please note that, in contrast to real-world applications, we are not calculating the metrics based on the observed outcome of the specific bet, but calculate the expected value for net win and RPS based on the true probabilities.

Table 1: Example calculations for model profitability and accuracy. Abbreviations: probA, probB, probabilities estimated by the forecaster; oddsA, oddsB, odds published by the bookmaker; expA, expB, expected net win of a bet when considering the forecaster probabilities; betA, betB, indicator of whether a bet was placed; $\widetilde{probA}$, $\widetilde{probB}$, *true probabilities; expWin, expected net win based on bets and true probabilities; expRPS, expected rank probability score based on forecasted and true probabilities.*

$probA$	$probB$	$oddsA$	$oddsB$	$expA$	$expB$	$betA$	$betB$	$\widetilde{probA}$	$\widetilde{probB}$	$expWin$	$expRPS$
40%	60%	1.9	1.9	-0.24	0.14	0	1	60%	40%	-0.24	0.28
50%	50%	1.9	1.9	-0.05	-0.05	0	0	60%	40%	0.00	0.25
60%	40%	1.9	1.9	0.14	-0.24	1	0	60%	40%	0.14	0.24

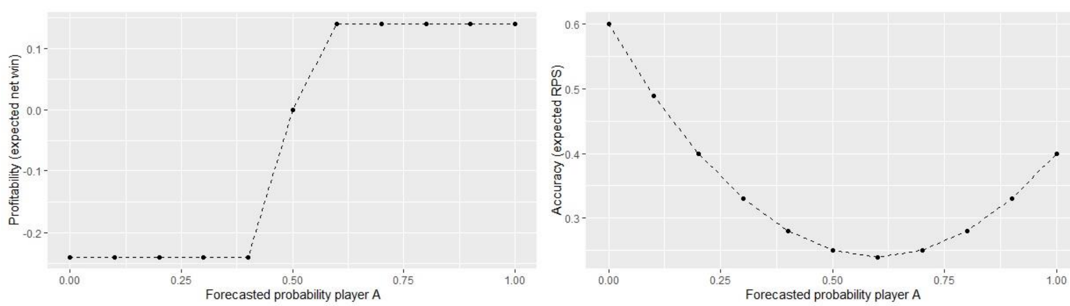


Figure 1. Illustration of profitability and accuracy when varying the probability estimation for player A in the forecasting model.

At the same time, we cannot assume that accuracy and profitability do not have anything in common. As an equally simple, but intuitive example, we assume that the forecasting model is perfectly accurate, not in a sense of predicting the winner, but in a sense of exactly estimating the true probabilities. In this case, the bettor would never face a systematic loss, i.e. in the long run, it is impossible for such a perfectly accurate model to be unprofitable.

Obviously, both examples are highly simplifying and constructed, however, they capture the main reasons why accuracy and profitability are different, though somehow related concepts in forecasting. This notion is supported by more complex investigations in various economic fields [12–15] including sports betting [10, 11].

2.2 Measures of profitability limit sample sizes

Measures of profitability (i.e. betting returns) are naturally subject to smaller sample sizes than measures of accuracy. The reason is that in many cases, the estimation of model and market may be so similar that none of the sides of a market promises profitability. Only in those cases where model and market are sufficiently in conflict, bets will be placed and even in these cases not on all sides. This issue is aggravated if bookmaker margins are high or if placing bets is restricted to those exceeding a specific expected net win as both implies that the difference between model and market needs to be high in order to make betting reasonable. As such, the sample size that profitability measures are based on is typically way lower than the sample size for accuracy, which implies higher noise in the results. Figure 2 illustrates this sample size problem based on simulated forecasts and bets. In each configuration, 10 million betting opportunities are simulated, where the bookmaker and the bettor are subject to a forecasting error in relation to the true probability. The bookmaker margin can be varied as well as a threshold representing the minimum expected net win of a betting opportunity that leads to a placed bet. Moreover, the correlation of the forecasting errors can be specified, but in this example is set to zero, i.e. uncorrelated errors are assumed. For each specification, the number of bets actually taken as a percentage of all betting opportunities can be calculated. For more details on the method used we refer to [10].

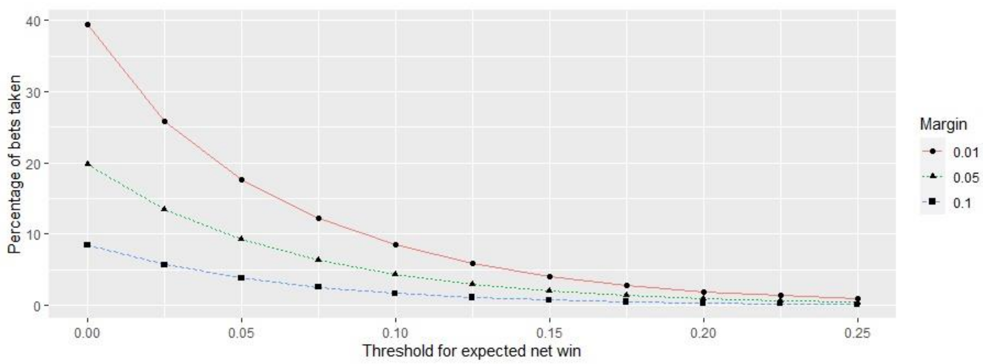


Figure 2. The percentage of bets taken for various bookmaker margins and various degrees of correlation between forecasting errors of model and bookmaker. Results are based on a simulation model with 10 000 000 betting opportunities each.

Results show that even for low margins and thresholds, the percentage of bets is limited and further decreases for increasing margins and increasing thresholds. For example, the sample size reduces to less than 10% of all betting opportunities if placing all bets with an expected net win above 5% and a margin of 5%. Like in every simulation, the process is simplified and the results depend on the assumptions of the simulation, particularly with regard to the degree and type of forecasting error made by both sides. However, similar sample size issues can be observed in the literature. For example, in the well-known ELO_goals model to forecast home, draw and away probabilities in football by Hvattum & Arntzen [16], the authors base accuracy measures on 15 181 matches, which equals 45 543 betting opportunities, of which only 87 (0.2%) to 12 152 (26.7%) are actually bet on, depending on the threshold used.

2.3 Profitability is usually not considered in model specification

When using a forecasting model to identify profitable bets, it is highly questionable whether the model itself has actually been optimized and selected to do so. Once a statistical model is designed, its parameters still need to be estimated in order to ensure a best possible fit to real-world data. However, this usually means that it is explicitly or implicitly fitted to the data in terms of accuracy by maximizing likelihood, minimizing least squared errors or RPS. For example, if using well-established statistical approaches like a regression model [16], the dependent variable is the observed outcome of matches. The respective betting odds are unknown to the regression model (and only used later in out-of-sample profitability) and thus the model cannot detect a model capability to exploit market biases. If using more sophisticated Machine Learning approaches, model specification becomes more complex than estimating a range of predefined parameters, however the basic issue prevails. As long as the model has been trained without knowledge of the betting odds, it is unlikely to capture specific options for profitability.

Despite this consideration, there is theoretical work [17] stating that under certain conditions the most accurate model (in terms of likelihood) is likewise the most profitable (in terms of a Kelly betting strategy). The mathematical derivation for this result, however, is based on assumptions that are usually not fulfilled in real-world betting.

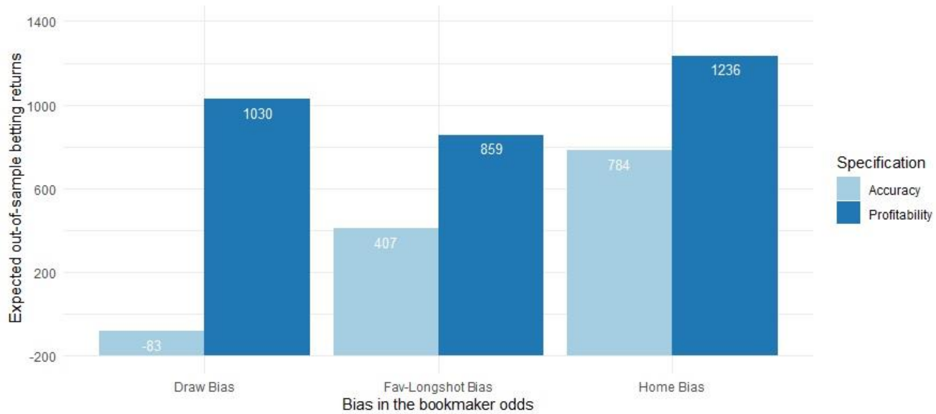


Figure 3. Out-of-sample profitability for models that were fitted for either accuracy (likelihood) or profitability (Kelly betting returns) in-sample. Results are based on three different biases, where the bookmaker slightly misestimates the draw occurrence, the favorite probability or the home advantage. Please note that the out-of-sample profitability refers to expected betting returns based on the ground truth in the simulation model and as such no error indicators need to be given.

For this reason, we compared the out-of-sample performance of models fitted for profitability (i.e. choosing the model specification with the highest betting returns) to models fitted for accuracy (i.e. the model specification with highest likelihood) in-sample. To do this, we use artificial data generated by a simulation framework [18] capable of simulating a ground truth on team ratings, outcome probabilities and bookmaker biases as well as applying forecasters models to these data and assessing their performance. Figure 3 illustrates the results of simulations based on 152.000 football matches, which corresponds to the number of games in 40 leagues over 10 seasons. A bookmaker margin of 10% and a Kelly betting strategy based on a fixed bankroll are used. Please note that the main findings do not change if changing the margin or applying a fixed stake betting strategy [19]. Results show that independent of the bookmaker bias, the models fitted for profitability yield higher expected out-of-sample profitability, being in line with our reasoning above. First attempts to consider betting odds in the model selection, e.g. by decorrelating bookmaker and model forecasts, have already been successfully pursued in the literature [6].

3 Discussion and conclusions

The present paper presented three issues in profitable sports forecasting, that should be more carefully considered in future studies. First, we are missing a clear distinction between profitability and accuracy models in sports forecasting. Only in a few exceptions [6, 20], the authors offensively state the emphasis on model profitability and it would be desirable if authors would be more clear in this regard. If the focus is on profitability, the accuracy should be of secondary importance, if at all and profitability might already be considered in the modelling process. If accuracy is the target, this should be clearly stated by authors and also acknowledged by editors and reviewers by not standardly asking for betting returns to be reported in the review process. Second, we are missing a stricter discussion on the inherent noise in betting returns. Authors can acknowledge and quantify this randomness in betting returns by using comparison to random betting strategies [10, 20] or bootstrapping methods [21]. Third, we are missing established ways to fit models for profitability. If parameter estimation for models is based on fitting the data for accuracy, the most profitable model specification might be systematically overlooked. As such, researchers would be well advised to already consider betting odds when selecting model parameters in-sample [11].

The present suggestions, particular on considering profitability for model selection, however, are not free from disadvantages. Particularly as the sample size problem of betting strategies might be extended to the in-sample model fitting. Moreover, fitting for profitability may requires approaches moving away from standard and established statistical concepts (like regression models). As such, it appears valuable to stimulate discussion on optimal ways to deal with the issues of model selection and evaluation in profitable forecasting.

Acknowledgements

This research was supported by a grant from the German Research Council (DFG, Deutsche Forschungsgemeinschaft) to D.M. Grant ME 2678/29.1

References

[1] Maher MJ (1982) Modelling association football scores. Statistica Neerlandica 36(3), 109–118.
[2] Wunderlich F, Memmert D (2021) Forecasting the outcomes of sports events: A review. European journal of sport science 21(7), 944–957. 10.1080/17461391.2020.1793002.
[3] Angelini G, Angelis L de (2019) Efficiency of online football betting markets. International Journal of Forecasting 35(2), 712–721. 10.1016/j.ijforecast.2018.07.008.
[4] Peeters T (2018) Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer results. International Journal of Forecasting 34(1), 17–29. 10.1016/j.ijforecast.2017.08.002.
[5] Pachur T, Biele G (2007) Forecasting from ignorance: the use and usefulness of recognition in lay predictions of sports events. Acta psychologica 125(1), 99–116. 10.1016/j.actpsy.2006.07.002.
[6] Hubáček O, Šourek G, Železný F (2019) Exploiting sports-betting market using machine learning. International Journal of Forecasting 35(2), 783–796. 10.1016/j.ijforecast.2019.01.001.
[7] Štrumbelj E, Vračar P (2012) Simulating a basketball match with a homogeneous Markov model and forecasting the outcome. International Journal of Forecasting 28(2), 532–542. 10.1016/j.ijforecast.2011.01.004.
[8] Koopman SJ, Lit R (2015) A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. J. R. Stat. Soc. A 178(1), 167–186. 10.1111/rssa.12042.
[9] Barnett T, O'Shaughnessy D, Bedford A (2011) Predicting a tennis match in progress for sports multimedia. OR Insight 24(3), 190–204. 10.1057/ori.2011.7.
[10] Wunderlich F, Memmert D (2020) Are betting returns a useful measure of accuracy in (sports) forecasting? International Journal of Forecasting 36(2), 713–722. 10.1016/j.ijforecast.2019.08.009.
[11] Hubáček O, Šír G (2023) Beating the market with a bad predictive model. International Journal of Forecasting 39(2), 691–719. 10.1016/j.ijforecast.2022.02.001.
[12] Boothe P, Glassman D (1987) Comparing exchange rate forecasting models. International Journal of Forecasting 3(1), 65–79. 10.1016/0169-2070(87)90079-3.
[13] Ertimur Y, Sunder J, Sunder SV (2007) Measure for Measure: The Relation between Forecast Accuracy and Recommendation Profitability of Analysts. J Accounting Res 45(3), 567–606. 10.1111/j.1475-679X.2007.00244.x.
[14] Fuertes A-M, Kalotychou E, Todorovic N (2015) Daily volume, intraday and overnight returns for volatility prediction: profitability or accuracy? Rev Quant Finan Acc 45(2), 251–278. 10.1007/s11156-014-0436-6.
[15] Leitch G, Tanner JE (1991) Economic forecast evaluation: profits versus the conventual error measures. The American Economic Review, 580–590.
[16] Hvattum LM, Arntzen H (2010) Using ELO ratings for match result prediction in association football. International Journal of Forecasting 26(3), 460–470. 10.1016/j.ijforecast.2009.10.002.
[17] Johnstone DJ (2011) Economic Interpretation of Probabilities Estimated by Maximum Likelihood or Score. Management Science 57(2), 308–314. 10.1287/mnsc.1100.1272.

[18] Garnica-Caparrós M, Memmert D, Wunderlich F (2022) Artificial data in sports forecasting: a simulation framework for analysing predictive models in sports. Inf Syst E-Bus Manage 20(3), 551–580. 10.1007/s10257-022-00560-9.
[19] Wunderlich F, Garnica-Caparrós M, Memmert D (2023) Profitable forecasting requires different model selection than accurate forecasting. Theoretical evidence from the domain of sports forecasting. Manuscript submitted for publication.
[20] Wheatcroft E (2020) A profitable model for predicting the over/under market in football. International Journal of Forecasting 36(3), 916–932. 10.1016/j.ijforecast.2019.11.001.
[21] Boshnakov G, Kharrat T, McHale IG (2017) A bivariate Weibull count model for forecasting association football scores. International Journal of Forecasting 33(2), 458–466. 10.1016/j.ijforecast.2016.11.006.

Learning and planning to solve boulder climbing problems

Mert Türedioğlu*, Barbaros Yet*, Kerem Demirtaş**

* Department of Cognitive Science, Graduate School of Informatics, Middle East Technical University, Ankara, Turkey, email address: byet@metu.edu.tr

** Smart Kiwi, London, UK

Abstract

Bouldering is a well-balanced sport between physical, technical and mental skill requirements as it tests decision making and problem-solving abilities of athletes alongside their physical fitness. In bouldering competitions, each climbing route is a novel decision problem that an athlete needs to solve in a limited amount of time. The athlete needs to plan the sequence of movements for a route they see for the first time before starting to climb. This study presents a goal-based climbing agent that learns from the previous solutions it observes to plan the sequence of actions for novel bouldering problems it encounters. The agent learns the cost of their movements from a dataset of the difficulty estimations and solutions provided by expert climbers and uses this information to solve the novel climbing problems it encounters. It also uses expert knowledge to constrain the learning and state space.

1 Introduction

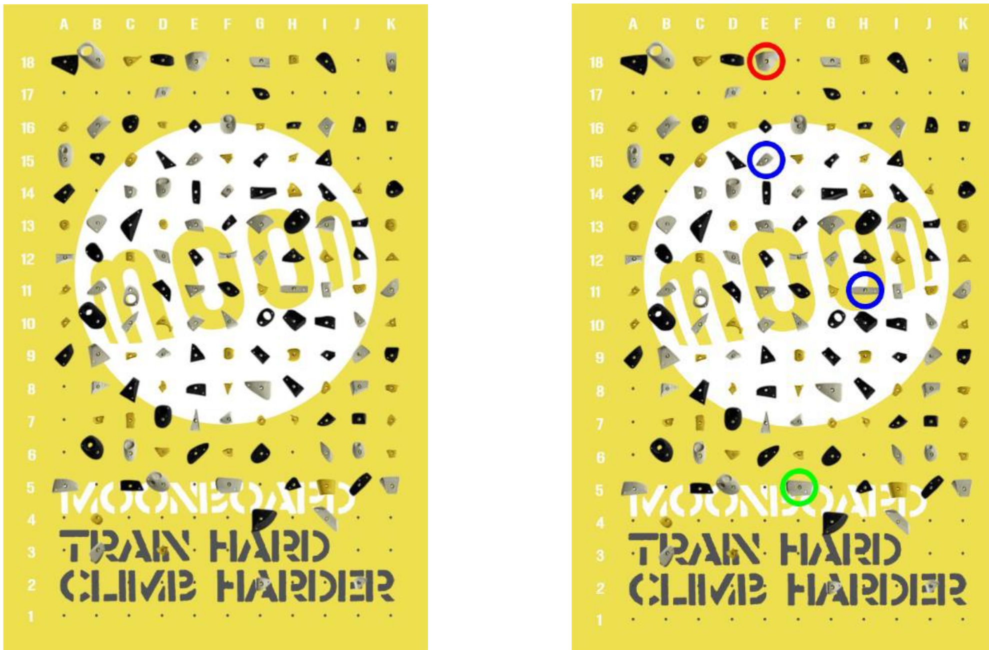
Bouldering is a challenging sport in terms of both physical and mental skills of athletes [1]. In order to successfully complete a climbing route, a climber plans and executes a series of movements in a limited amount of time. Each route has unique properties hence a plan applies only to the specific route being attempted by the climber. A solution to a climbing route is called a beta. An inferior beta increases the risk of failure and exhaustion. Climbers use indoor training boards to train their physical and problem-solving skills. A typical training board is composed of holds of different kinds on a single plane. Some training boards are standardized to provide a shared training medium to the climbing community. Climbers can create and share new climbing problems for standardized training boards and discuss their solutions.

This paper proposes a novel approach to solve climbing problems on standardized training boards. The proposed approach tackles the climbing as an optimization problem that aims to minimize the total cost of movements. The cost of a movement is not equivalent to its distance. We learn the cost of movements from a dataset of climbing problems and their difficulty grades. We focus on Moonboard which is one of the most popular standardized training boards used by a large international community of climbers. Moonboard users share the climbing problems they created and graded through a widely used online application. The standardized layout of Moonboard and the problems shared by other athletes in the online application facilitates digitalized data collection for climbing problems.

In the remainder of this paper, Section 2 gives an overview of Moonboard and reviews previous computational studies applied to training boards. Sections 3 and 4 present our approach and the results of applying it to climbing problems. Section 5 presents our conclusions.

2 Moonboard climbing

Moonboard has a standard grid structure on a 40-degree angled plane. The grid structure is composed of the main board that is 18 x 11 grid and the kicker that is 5 x 2 grid. The original 2016 layout of Moonboard has been revised in 2017 and 2019. All revisions have the same kicker grid but the layout of the holds on the main board differs. This study uses the original 2016 layout which has 140 holds on the main board that is shown in Figure 1a. The kicker is not shown in this figure.



a) b)

Figure 1. Moonboard a) 2016 Layout b) Example Route

Figure 1b shows a climbing problem example [2]. The climber is allowed to grip the holds highlighted by circles. Green circles indicate the starting holds, blue circles indicate the intermediate nodes, and the red circle indicates the target hold. The athlete starts climbing with both hands touching the staring holds. The climber's feet can touch any kicker node throughout the climbing process. The goal is to grasp the target hold with two hands.

Climbing problems have standard grading scales indicating their difficulty. The Fontainebleau grading system and the Hueco scale are commonly used grading scales for bouldering problems. Different grade scales can be converted to each other. There is no formal definition or algorithm to grade a climbing problem. It is defined based on domain knowledge by comparing it to other similar climbing problems. The grades of problems can change by a few grades over time in consequent reviews as some of the neighbor grades are similar to each other. In order to reduce the number of grades for our learning task and cover the possible changes in grades we used a simplified grade that merges neighbor grades with similar difficulties. Table 1 shows those grading scales and their conversions.

The moderators of the Moonboard application review the content and grades of climbing problems uploaded by the users. The problems that have been confirmed by the moderators are called benchmark problems. This review is required because any user can upload a climbing problem to the application, and some of the problems can be physically impossible or inaccurately graded.

Table 1. Bouldering Grade Scales

Hueco	Fontainebleau	Simplified	Hueco	Fontainebleau	Simplified
VB	3	Basic	V7	7A+	Advanced
V0-	4-		V8	7B	
V0	4			7B+	
V0+	4+		V9	7C	
V1	5		V10	7C+	Elite
V2	5+		V11	8A	
V3	6A		V12	8A+	
	6A+		V13	8B	
V4	6B		V14	8B+	
	6B+	Intermediate	V15	8C	
V5	6C		V16	8C+	
	6C+		V17	9A	
V6	7A				

The standard grid structure of Moonboard can easily be transformed into machine readable data hence it is well suited for computational and modelling studies. Previous computational studies mainly focused on grade estimation and new route generation. Dobles et al. [3] compared various machine learning methods for grading Moonboard problems. Tai et al. [4] graded Moonboard problems using graph convolutional neural networks. Duh and Chang [5] graded Moonboard problems by first creating a solution for the problem using Recurrent Neural Networks. Çelik [6] used a rule-based model to compute the difficulty of a route. Stapel [7] used a heuristic approach for creating climbing routes. Seal and Seal [8] examine the difficulty and physical requirements of climbing tasks by creating random artificial training boards and analyzing the shortest path between the start and end holds of those boards.

3 Proposed approach

In this section, we present our proposed approach to solve Moonboard climbing problems. The solutions of our approach provide a full 'beta' involving the sequence of movements for both hands and feet. We tackle climbing as an optimization problem in which the climber needs to find the sequence of movements with the minimal total cost. The cost of a movement is not equivalent to the distance of the target hold, it is learned from data and depends on other factors than distance such as the limbs that are not gripping any hold.

We represent the problem as a hypergraph in which hypernodes represent the state of the climber, and hyperedges represent the actions that change the state. A hypernode is in the form of ["Left-Hand", "Right-Hand", "Left-Foot", "Right-Foot]. For example, [G8, H11, None, K5] represents a climber who has her left hand, right hand and right foot on G8, H11 and K5 respectively, and her left foot does not touch any of the holds. The possible actions of a climber change the state corresponding to her hands or feet. For example, moving the left-hand from G8 to E13 in this state is represented as a hyperedge which connects this hypernode to [E13, H11, None, K5]. If we include hypernodes for all combinations of states and hyperedges for all possible transitions from a state, the size of the hypergraph can get very large. However, many of those hypernodes and hyperedges representing states and actions are clearly impossible for a human climber. For example, a climber cannot be in an upside-down state such as [G8, H11, G16, B16] because of the very steep angle of the board and the size of the holds. Similarly, a climber cannot move from [G8, H11, None, K5] to [A15, H11, None, K5] because of her arm-length and center of gravity. We filter those infeasible nodes to reduce the size of the graph. The heuristics we use for filtering out infeasible nodes and edges include removing nodes beyond the physical capabilities of humans due to height and arm-span, removing extreme actions such as 3-meter jumps or a state change for all four limbs.

Actions have different difficulties for a climber. Some moves have a higher risk of failure and falling. The difficulties of different moves are represented as the costs associated with the corresponding hyperedges in the hypergraph. The difficulty depends on the distance of how the center point of the climber changes. If a climber moves to a further hold which makes a big change to her center of gravity, there is more risk of failing the move. The number of free limbs also increases the difficulty of a particular move. For example, if a climber has one of her feet or hand free, this may increase the risk of failure for the next move. We learned the move difficulties from a dataset of climbing routes and solutions. The difficulties of each move were labelled with the grade of the associated climbing route. We trained a Bayesian ordinal regression model that predicted the difficulty grade of a move for a given action. The Bayesian model used the distance in center point and the number of free limbs associated with the move for this prediction. Using this model, we assigned costs to hyperedges. After the hypergraph is prepared, filtered, and the costs of edges are assigned, the optimal climbing plan is found by the Dijkstra's shortest path algorithm.

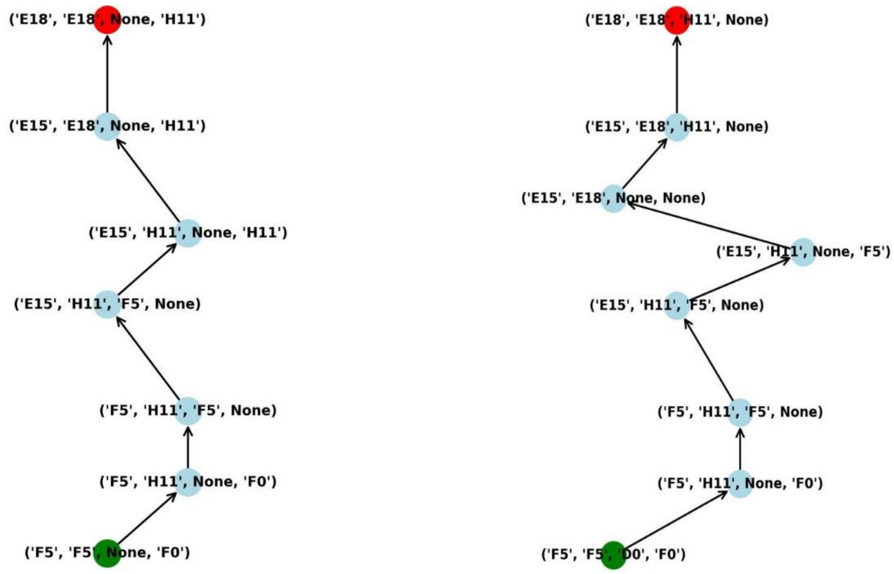
4 Results

Figure 2a shows the solution of our approach for the example problem shown in Figure 1b. The green node in Figure 2 is the starting condition [F5, F5, None, F0] indicating the climber starts with both hands on F5, left foot free and right foot touching a hold in the kicker part of the board. The sequence of hand and feet movements to reach the goal condition can be read from this graph. For example, the first move is to move the right hand from F5 to H11. The second move is to move the left feet to F5 and free the right foot. The proposed approach solves this problem in 6 moves.

Figure 2b shows the suggested solution for the same problem in the Moonboard application. Table 2 compares our solution with the suggested solution. All hand moves in our solution were in accordance with the suggested solution. The suggested solution involves one more footstep. Our solution has a smaller total distance than the suggested solution. Our solution is suitable for shorter climbers as the average distance of the moves is smaller than the suggested solution. Moreover, the moves in our solution requires less power compared to the suggested solution. In the suggested solution, the move ['E15, 'H11', None, 'F5'] to ['E15, 'E18', None, None] requires large amount of power and is difficult for shorter climbers.

Table 2. Bouldering Grade Scales

Problem ID: 50356					
Proposed Approach			*Suggested Solution*		
# moves	# hand moves	# foot moves	# moves	# hand moves	# foot moves
6	4	2	7	4	3
total distance	mean distance		total distance	mean distance	
594.4	99.1		776.2	110.9	



a) b)

Figure 2. Solution a) provided by the proposed approach b) suggested in the Moonboard application

Apart from this case, we compared the solutions of our approach with the suggested solutions for 30 benchmark routes from the Moonboard application. Our approach made the same amount of hand moves as the suggested solutions for benchmark routes. Our approach made 6.2 hand moves on average, the suggestion solutions had an average of 6.26 hand moves. Our approach made fewer amount of feet moves compared to the suggested solutions. On average, our approach made 2.3 foot moves while the suggested solutions had 6 foot moves. In some cases, smaller number of foot moves can be advantageous for the climber. However, fewer foot moves often cause inefficient body positioning and increase the strength requirements. Both our approach and proposed solutions tend to use most holds available in a problem. On average, our approach and the suggested solutions used 88% and 91% of the holds respectively.

5 Conclusion

This paper described an approach to provide solutions to climbing problems on a training board for bouldering called Moonboard. The proposed approach learns the cost of hand and foot movements from a dataset of climbing problems and their difficulty grades, and computes the shortest path solution based on those costs. The solution includes the sequence of hand and foot movements required to reach the goal state of the problem. We compared the solutions provided by our approach to the solutions suggested by expert climbers in the Moonboard application. Both solutions were similar in terms of their hand moves. The solutions of our approach tend to have less foot movements and shorter moves compared to the suggested solutions by expert climbers. The suggested solutions are not necessarily globally optimal as they depend on fitness and anthropometric properties of the expert climber providing it. In future studies, we plan to revise and validate the model by asking climbers to follow the advice of our model and provide feedback accordingly.

References

[1] Hörst, E. J. (2008). *Training for climbing : the definitive guide to improving your performance*. Falcon Guides.
[2] Moonboard (2023) Moonboard problem – The Crane Direct [Accessed: 31/05/2023] https://www.moonboard.com/Problems/View/50356/the-crane-direct
[3] Dobles, A., Sarmiento, J. C., & Satterthwaite, P. (2017). Machine Learning Methods for Climbing Route.
[4] Tai, C.-H., Wu, A., & Hinojosa, R. (2020). Graph Neural Networks in Classifying Rock Climbing Difficulties.
[5] Duh, Y.-S., & Chang, R. (2021). Recurrent Neural Network for MoonBoard Climbing Route Classification and Generation. https://doi.org/10.48550/arXiv.2102.01788
[6] Çelik, R. (2022). *Model-Based Difficulty Estimation of Indoor Bouldering Routes*. Utrecht University.
[7] Stapel, F. (2020). A Heuristic Approach to Indoor Rock Climbing Route Generation.
[8] Seal, D., & Seal, R. (2022). *Optimum Route Computation in a Chaotic Artificial Climbing Wall* (pp. 671–680). https://doi.org/10.1007/978-981-16-5987-4_68

Does draft currency promote competitive balance? An empirical investigation of the National Football League 2002-2021

Michael A. Lapré* and Elizabeth M. Palazzolo**

* Owen Graduate School of Management, Vanderbilt University, Nashville, TN, USA, m.lapre@vanderbilt.edu
** Lazard, New York, NY, USA, elizabeth.palazzolo@lazard.com

Abstract

In the National Football League (NFL) annual draft, teams take turns selecting entering players. The draft is a market mechanism designed to promote competitive balance as the NFL assigns draft positions to teams in reverse order relative to last season's performance. Teams frequently trade draft picks for other picks and/or players. We use several market valuations of draft picks to define original draft currency as the total value of draft picks available before any trades and final draft currency as the total value of picks used after all trades. For the 2002-2021 period, counter to the intent of the draft, we find that original draft currency does not affect the probability of reaching the playoffs, but final draft currency does. Usage of outdated market valuations of draft picks by most teams can help explain how the best team has used draft-pick trades to remain among the strongest teams over two decades thereby perpetuating competitive imbalance.

1 Introduction

Competitive excitement, in particular the uncertainty of the outcome of a game and of a season, attracts fans (Fort 2012). Competitive balance, meaning teams of equal strength, boosts outcome uncertainty. The National Football League (NFL) uses a draft for entering players ("rookies") as a market mechanism to promote competitive balance. After the conclusion of a season, the NFL creates a draft order assigning positions to teams in reverse order relative to last season's performance. The team with the worst win-loss record gets the first position, and the Super Bowl champion gets the last position. The draft proceeds through seven rounds. In each round, each team owns the right to pick one player at their assigned position. The idea is to promote competitive balance as the worst team can pick the best available player. Drafted players can only sign a rookie contract with the team that selected them. After a rookie contract expires, the player becomes a free agent and can sign with any team. Teams can trade their assigned draft picks for other picks and/or players. From 2002 to 2021, only 61% of all draft picks were used by teams according to their assigned draft position, whereas 39% of draft picks were traded.

The NFL draft's goal of promoting competitive balance remains empirically ambiguous (Maxcy 2012). Few prior papers have studied whether the draft improves competitive balance. Grier and Tollison (1994) find that a later average draft position in the last 5, 4, and 3 years has a negative relation with a team's win percentage controlling for last year's win percentage. Similarly, Caporale and Collier (2015) find that a later average draft position over the last 5 years has a negative relation with a team's win percentage controlling for last year's win percentage and strength of schedule computed with a team's opponents win percentages from last year. While these findings are promising, there are several issues with these studies. First, win percentage is a problematic dependent variable as teams have different schedules facing opponents with different strengths. The standings at the end of the regular season determine which teams advance to the playoffs–the single-elimination tournament that culminates with the Super Bowl to determine the NFL champion. Teams might temporarily improve win percentage yet still miss the playoffs. Second, there might be a performance spillover effect longer than one year ago. Third, the models assume equal contribution from each draft year. Fourth, using draft position assumes that the value of draft picks follows a linear decreasing function of draft position. Fifth, the models only consider original draft position ignoring draft-pick trades. We address all five issues in this paper.

In this paper, we investigate whether the NFL draft promotes competitive balance. We define (i) original draft currency as the total value of picks available before any draft trades, and (ii) final draft currency as the total value of picks used after all draft trades. We address the following questions. Does higher original draft currency increase the probability of reaching the playoffs? If so, after how many years? How does final draft currency affect the probability of reaching the playoffs? What market valuation of draft picks best captures any reduction in competitive imbalance?

2 Market Value for NFL Draft Picks

In 1991, Mike McCoy, a minority owner of the Dallas Cowboys, used a subset of prior draft-pick trades to assign values to each draft pick (Massey and Thaler 2013). His objective was to characterize past trading behavior rather than create a price list. However, McCoy's draft pick value chart (PVC) would revolutionize the NFL draft. Initially, the Dallas Cowboys used the PVC in the early 1990s. Subsequently, personnel from the Dallas Cowboys moved to other teams taking the PVC with them. Soon the PVC become the norm for negotiating draft-pick trades. In 2004, ESPN reported that "Before NFL general managers consider trading draft picks, they more often than not consult this value chart." (ESPN Insider 2004). In Figure 1, the PVC graph is rescaled relative to the number 1 pick.

Massey and Thaler (2013) empirically estimate the value of draft picks using all trades involving only same-year draft picks. Graph MT in Figure 1 is Massey and Thaler's (2013) estimated draft position value curve. MT resembles PVC indicating that teams did indeed use PVC for negotiating draft-pick trades. The founder of Pro Football Reference (PFR), Doug Drinen, created the Approximate Value (AV) method to assign a single number to the on-field value of a player (Winston et al. 2022). Using draft data from 1970 to 1999, Stuart (2008) estimated a formula linking career AV of each player to his rookie draft slot. PFRAV in Figure 1 is Stuart's draft-value curve rescaled relative to the number 1 pick. After analyzing games played, games started, career AV, and Pro Bowl (an all-star game played toward the end of the season) starts, Schuckers (2011) used the drafts from 1991 through 2001 to propose an Alternative Pick Value Chart. APVC in Figure 1 is Schuckers' chart rescaled relative to the number 1 pick. Next, we investigate our research questions for each of the curves in Figure 1.

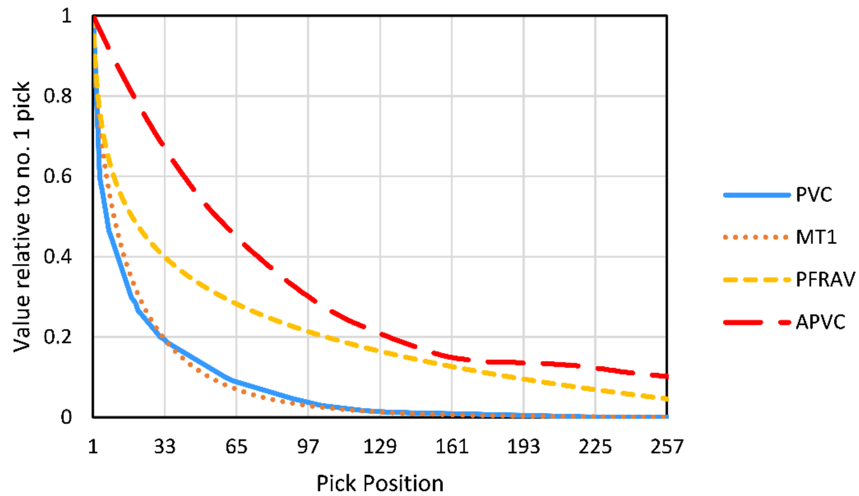


Figure 1. Draft-value curves.

3 Empirical Analysis

We collected data on the NFL drafts and all NFL games played for the 1998 through 2021 seasons from Wikipedia and pro-football-reference.com. NFL teams are assigned to either the American Football Conference (AFC) or the National Football Conference (NFC). In 2002, the NFL expanded to 32 teams and in both conferences realigned to four divisions (East, North, South, and West) of four teams each. From 2002 onwards, in each conference, all four division winners qualified for the playoffs as well as two wild-card teams (non-division winners with the best record). So, 12 out of 32 NFL teams qualify for the playoffs. Starting in 2021, a third wild-card team also qualified in each conference. The 2002 realignment provides an opportunity to study the same set of 32 teams. Thus, we focus on the 2002-2021 period.

Our dependent variable is $PO_{it} = 1$ if team *i* qualified for the playoffs at the conclusion of regular season *t*, 0 otherwise. Let OS_{it} be the original set of draft picks for team *i* in year *t* before any draft trades. Similarly, with FS_{it} we denote the final set of draft picks used by team *i* in year *t* after all draft trades. We use $v(n)$ to denote the value of the *n*-th draft pick. We calculate the original draft currency for team *i* in year *t* as $ODC_{it} = \sum_{n \in OS_{it}} v(n)$. Analogously, we calculate final draft currency as $FDC_{it} = \sum_{n \in FS_{it}} v(n)$.

To control for the strength of the teams each season, we use least squares ratings. Because each NFL team plays a different subset of opponents, least squares ratings are better control variables for team strength compared to win percentages. Let r_{it} be the rating for team *i* in season *t*. The rating represents the expected scoring margin on a neutral field against an average team in season *t*. The scoring margin m_{ijt} is the actual difference between the number of points scored by home team *i* and the number of points scored by away team *j* in season *t*. Let h_t be the home advantage in season *t* and N_t the number of teams in season *t*. For 2002 through 2021, $N_t = 32$. Lastly, let $G(t)$ be the set of all regular season games played in season *t*.

The forecasted margin for a game between home team *i* and away team *j* in season *t* is $h_t + r_{it} - r_{jt}$. Consequently, the forecasted error for this game is $(h_t + r_{it} - r_{jt}) - m_{ijt}$. For each regular season *t*, we solve for the team ratings (r_{it}) and the home advantage (h_t) by minimizing the sum of squared errors subject to the average rating being equal to zero:

$$\min \sum_{(i,j)\in G(t)} ((h_t + r_{it} - r_{jt}) - m_{ijt})^2$$

$$\text{s.t. } \frac{1}{N_t}\sum_{i=1}^{N_t} r_{it} = 0$$

Next, we calculate the strength of schedule SoS_{it} by averaging the ratings of the opponents faced by team i in all regular season games in season t. We use logistic regression to estimate the probability of reaching the playoffs as follows:

$$\ln\frac{\Pr(PO_{it}=1)}{1-\Pr(PO_{it}=1)} = \alpha_0 + \alpha_1 r_{it-1} + \alpha_2 r_{it-2} + \alpha_3 r_{it-3} + \alpha_4 r_{it-4} + \alpha_5 SoS_{it}$$
$$+\beta_1 ODC_{it} + \beta_2 ODC_{it-1} + \beta_3 ODC_{it-2} + \beta_4 ODC_{it-3} + e_{it} \quad (1)$$

We estimate Equation (1) for each of the draft-value curves in Figure 1, i.e., for each curve, we use the plotted $v(n)$ values in Figure 1 to calculate original draft currency. To assess the impact of final draft currency, we replace the original draft currency variables with final draft currency variables in Equation (1). In the estimation, we fit random-effects models adjusting the standard errors for clustering observations by team.

Table 1. Logistic Regression Models for Making the Playoffs in 2002-2021: Original Draft Currency

Original draft currency	(1) -	(2) PVC	(3) MT	(4) PFRAV	(5) APVC
Rating, lag 1	0.088***	0.100***	0.099***	0.110***	0.109***
Rating, lag 2	0.063***	0.085***	0.088***	0.052**	0.048**
Rating, lag 3	–0.019	–0.004	–0.006	–0.016	–0.022
Rating, lag 4	0.017	0.008	0.006	0.029	0.032
Strength of schedule	–0.319***	–0.326***	–0.325***	–0.318***	–0.317***
Draft currency		0.446	0.373	0.675	0.552
Draft currency, lag 1		0.726	0.759	–0.148	–0.240
Draft currency, lag 2		0.575	0.468	0.162	–0.037
Draft currency, lag 3		–0.216	–0.236	0.453	0.432
Constant	–0.580***	–1.548	–1.451	–2.553	–2.396
Log pseudolikelihood	–370.90	–369.76	–369.81	–369.47	–369.16

Dependent variable: Reach the Playoffs. * Significant at 0.05, ** at 0.01, and *** at 0.001.
Number of observations: 635.

Table 1 shows the logistic regression estimates for reaching the playoffs with original draft currency. Model (1) is the base model with control variables only. The positive and statistically significant coefficients for ratings from the past two years indicate that higher team strength in the past two years enhances the probability of reaching the playoffs. Further back in time, the effect of team strength drops off. Ratings from three or four years ago are not statistically significant. The negative and statistically significant coefficient for strength of schedule indicates that playing tougher opponents reduces the probability of reaching the playoffs. These significance findings are consistent across all models in Table 1. In Model (2), we add the original draft currency variables calculated with $v(n)$ values for the PVC. None of the draft currency variables are statistically significant. Thus, higher original draft currency measured with PVC does not lead to a higher probability of reaching the playoffs. We find the same in Models (3) through (5) for MT, PFRAV, and APVC.

Table 2. Logistic Regression Models for Making the Playoffs in 2002-2021: Final Draft Currency

	(1)	(2)	(3)	(4)	(5)
Final draft currency	-	PVC	MT	PFRAV	APVC
Rating, lag 1	0.088***	0.065***	0.063***	0.075***	0.076***
Rating, lag 2	0.063***	0.072**	0.072**	0.060***	0.060***
Rating, lag 3	–0.019	0.010	0.009	0.011	0.006
Rating, lag 4	0.017	0.035	0.036	0.033	0.030
Strength of schedule	–0.319***	–0.323***	–0.325***	–0.314***	–0.315***
Draft currency		–0.606	–0.616	–0.177	–0.130
Draft currency, lag 1		0.261	0.263	–0.174	–0.152
Draft currency, lag 2		0.916*	0.828*	0.664*	0.424*
Draft currency, lag 3		0.759	0.690	0.639*	0.448**
Constant	–0.580***	–1.437*	–1.341*	–2.241**	–2.109**
Log pseudolikelihood	–370.90	–365.67	–365.83	–362.95	–362.95

Dependent variable: Reach the Playoffs. * Significant at 0.05, ** at 0.01, and *** at 0.001.
Number of observations: 635.

Table 2 shows the logistic regression estimates for reaching the playoffs with final draft currency reflecting actual picks made after all draft trades. Again, the significant estimates for the control variables (prior team ratings two years back and strength of schedule) are consistent across all models. In Model (2), we add the final draft currency variables calculated with $v(n)$ values for the PVC. Final draft currency right before the current season as well as for the previous year is not significant. Yet, the coefficient for final draft currency from two years ago is positive and significant. So, draft picks made two years ago valued with the PVC increased the probability of reaching the playoffs. Final draft currency from three years ago is not significant. We find the same pattern in Model (3) when we calculate final draft currency with MT. In Models (4) and (5), both final draft currency from two years ago as well as three years ago are positive and statistically significant. Furthermore, the higher log pseudolikelihood values indicate better model fit for Models (4) and (5) compared to Models (1) through (3). These results suggest that draft-value curves PFRAV and APVC better capture how final draft currency from two and three years ago increases the probability of reaching the playoffs.

4 Conclusion

For each draft curve, we find that original draft currency does not affect the probability of reaching the playoffs. Interestingly, after a few years, final draft currency does affect access to the playoffs. Since the total amount of original draft currency is a fixed pie each year, trading picks is a zero-sum game among the 32 teams. Some teams gain draft currency at the expense of other teams. So, the draft does not promote competitive balance, but teams that "out-trade" other teams increase their chances of reaching the playoffs. It takes at least until the third season after a draft before the impact on access to the playoffs can be expected. Lastly, we find that model fit is better for draft curves with a less steep decline in draft-pick value (PFRAV and APVC) than the PVC.

During 2002-2021, the New England Patriots are by far the most successful team with 17 playoff appearances, 8 Super Bowl appearances, and 5 Super Bowl wins. The Patriots trade more of their draft picks (58%) compared to the league average (39%). As the strongest team, they had the least original draft currency. Trading picks they ended up with more final draft currency –measured by PVC– than 4 other teams. However, measured by APVC (which gives better model fit), the Patriots ended up with more final draft currency than

21 other teams. The Patriots acquired draft picks undervalued by other teams which allowed the Patriots to remain among the strongest teams for two decades thereby negating the intent of the draft. The mechanism of the draft is not necessarily flawed. Instead, other teams' perceived value of draft picks perpetuates competitive imbalance.

References

[1] Caporale, T. and Collier, T.C. (2015) *Are we getting better or are they getting worse? Draft position, strength of schedule, and competitive balance in the National Football League*. Journal of Labor Research **36**, 291-300.

[2] ESPN Insider. (2004, March 1). *NFL draft-pick value chart*. Retrieved from https://www.espn.com/nfl/draft06/news/story?id=2410670

[3] Fort, R. (2012) *Competitive balance in the NFL.* In *The Economics of the National Football League: The State of the Art.* (Quinn, K.G., Ed.), Springer, New York, pp.207-224.

[4] Grier, K.B. and Tollison, R.D. (1994) *The rookie draft and competitive balance: The case of professional football*. Journal of Economic Behavior and Organization **25**(2), 293-298.

[5] Massey, C. and Thaler, R.H. (2013) *The loser's curse: Decision making and market efficiency in the National Football League draft*. Management Science **59**(7), 1479-1495.

[6] Maxcy, J. (2012) *Economics of the NFL player entry draft system.* In *The Economics of the National Football League: The State of the Art.* (Quinn, K.G., Ed.), Springer, New York, pp.173-186.

[7] Schuckers, M. (2011) *An alternative to the NFL draft pick value chart based upon player performance*. Journal of Quantitative Analysis in Sports **7**(2), Article 10.

[8] Stuart, C. (2008) *The draft value chart: Right or wrong?* https://www.pro-football-reference.com/blog/index9a9b.html?p=527 (accessed May 5, 2022).

[9] Winston, W.L., Nestler, S., and Pelechrinis, K. (2022) *Mathletics* (2nd Ed.). Princeton University Press.

Evaluation of the performance of golfers on the LPGA tour based on LGCM

Beom-Jin Kim* and Taeho Kim**

* WCA Canada, 36 Dolmugae-Gil, Gwacheon,
Gyeonggi-Do, South Korea, Korea (ROK), email address: bkim25@wcacanada.com
** rm. 505, bldg. 14, Incheon National University, 119 Academy-ro, Yeonsu-gu, Incheon, Korea (ROK), email address: tacokim@gmail.com

Abstract

A latent growth curve model (LGCM) is used to examine the performance of golfers on the LPGA tour. The panel data come from the last three seasons (2020, 2021, and 2022). The dependent variable is the official money list of players, while the independent variables consist of efficiency factor, technical factor (the number of birdies, eagles, and holes in one), and mental factor (driving accuracy and putting average) of players. Each factor other than the efficiency factor is determined by the weighted combination of parenthesized variables using Principal component analysis (PCA), and is measured as two latent variables of the intercept and the slope of the growth pattern for 3 seasons. The efficiency factor is measured by using data envelopment analysis (DEA) and its bootstrapping with the output factors of scoring average and the percent of rounds in the 60s and the input factors of previous research. Harvey Penick said that golf performance is dependent on various factors, and that golf is psychology. These issues are thus investigated.

1 Introduction

The game we know as golf has been very popular since the mid-15th century, and saw noticeable growth during the COVID-19 pandemic. In Korea, the number of golfers increased by 16.4% between 2017 and 2021, a trend that is expected to continue (https://biz.chosun.com/topics/topics_social/2023/01/31/DPKERVJRE5AA7AYSDCLJEZJ6X4/). This growth in the popularity of golf has also resulted in more people wanting to become professional golfers. Notably, the average money list per LPGA (Ladies Professional Golf Association) event is greater than $3 million, with the best player in a year earning more than $4 million (https://www.lpga.com/).

However, golf is also known to be one of the hardest sports to master. One study ranked golf 51st in the list of toughest sports, and 5-7th in the ability to react quickly to sensory perception and the ability to evaluate and react appropriately to strategic situations (https://www.espn.com/espn/page2/sportSkills). This implies that both good shooting techniques and mental stability are essential to becoming a good golfer. In fact, both professional and amateur players show substantial variation in performance depending on their mental state.

This research examines the effect of technical factors, mental factors (Chung, et al., 2017), and efficiency (Fried, et al., 2004) on the official money list of professional golf players in LPGA. While previous research has looked only at the effect of one (Chung, 2020; Kwak and Kim, 2022) or a couple of factors (Chung, et al., 2017; Kim and Chae, 2021) on a golfer's performance in a season, this research focuses on the integrated effect of all the factors. Technical factors may indicate the ability to react quickly to sensory perception, mental factors may indicate the ability to evaluate and react appropriately to strategic situations, and technical efficiency may mean the ability of players to connect techniques in order to score. Additionally, we try to examine the growth of players in technical factors, mental factors, efficiency, and money list, together in 3 seasons. Though Kim and Min (2014) analyzed panel data on golfers, it is different in that each factor is analyzed separately.

In Sections 2 and 3, we introduce the research approaches and data. The results of our empirical analysis are presented and explained in Section 4. Finally, Section 5 summarizes the paper.

2 Research methods

This research uses 4 research methods. To integrate sub-factors for mental factor and technical factor Principal Component Analysis (PCA) is used, while Data Envelopment Analysis (DEA) is used to evaluate the technical efficiency of players and DEA bootstrapping is also used to obtain more robust efficiency scores. Finally, we use the Latent Growth Curve Model (LGCM) to estimate the relationship between independent factors (mentality, technique, and efficiency) and the money list, and to examine their growth from the 2020 season to the 2022 season.

Specifically, PCA coefficients which explain at least 90% of variance are taken to make an integrated variable for mentality and technique every season. Minitab 16 software uses PCA (Lind, 2021). DEA, which is a nonparametric approach to estimating the efficient frontier and efficiency scores, is implemented with 4 input factors and 2 output factors for each season, and a BCC model is used assuming variable returns to scale (Charnes et al., 1994). DEA bootstrapping was also run 2,000 times for each season to avoid sampling errors from the small sample size. Lingo 14 optimizer was used for DEA, and an app developed by the authors in Microsoft Visual Studio 2010 and integrated with Lingo 14 optimizer was used for bootstrapping. In LGCM, the observed variables are repeated measures of a variable and the latent variables are constructs representing patterns of change in a variable. Essentially, two latent variables are specified to represent the patterns of change: intercept and slope (Preacher et al., 2008). LGCM was designed and implemented by IBM AMOS 22 in this paper (Lee and Lim, 2015).

3 Data

Data comes from the top 60 players in the 2020 season money list ranking of LPGA. We extracted each value from the sit one by one from the LPGA official homepage (https://www.lpga.com/players). The reason why only 52 players were included is that 8 players were only included in one of the 2021 and 2020 season statistics. The input-output specifications for DEA came from Fried et al. (2004). Two technical factors are the number of birdies and the number of eagles (Chung et al., 2017); the number of hole-in-ones is omitted because of the lack of variation across players. For mental factors, while we ideally should have done a survey of the players on their psychological capital, this was impossible due to our lack of access to the LPGA players. We decided to find indirect measures reflecting the mentality of players and concluded that driving accuracy and putting accuracy are qualified (Penick et al., 2012).

Table 1 summarizes the variables in 2022 season data which are used for LGCM analysis, PCA, and DEA. Summaries of the 2021 and 2020 season data are not shown here due to a lack of space. As seen in the table, there are substantial variations in all variables, suggesting the need for further analyses.

Table 1. Summary of data of 2022 season

Variables		N	Min	Max	Mean	Standard deviation
D	Official Money List ($)	52	131,874	4,364,403	1,053,231.83	802,769.15
M	Driving Accuracy	52	0.64	0.86	0.75	0.05
M	Total # of putts per 18 holes (Putting accuracy)	52	28.61	31.67	29.87	0.55
T	Birdies	52	101	396	263.35	53.78
T	Eagles	52	0	11	5.60	3.19
E/X	Average Driving Distance (yds.)	52	240.64	279.25	258.29	9.29
E/X	Greens in Regulation	52	0.67	0.78	0.72	0.03
E/X	# of Putts per green hit in regulation	52	1.72	1.85	1.79	0.03
E/X	Sand Saves	52	0.30	0.66	0.49	0.07
X	Career (years)	52	3	16	8.81	3.74
E/Y	Scoring Average	52	68.99	71.93	70.62	0.68
E/Y	Rounds in the 60s	52	7	49	25.94	8.02

D: Dependent variable of LGCM, M: Mental factors of PCA, LGCM, T: Technical factor of PCA, LGCM, E/X: Input factor of DEA, X: Independent variable of LGCM, E/Y: Output variable of DEA

4 Result

As noted, the result of this paper consists of three parts: the efficiency scores of players, which implies the ability of players to connect their techniques to a final score; the PCA result to integrate two sub-factors into a factor from couple factors for mental factor and technical factor, respectively; and, the LGCM result, to find influencing factors on the money list representing the growth of players in variables.

4.1 Efficiency

To examine how 4 input factors are related to 2 output factors, DEA based on BCC was implemented and the result is illustrated in Table 2. As seen here, players have been very efficient and there is little variation among players. There is no growth or contraction through 3 seasons. The reason for this is that very high-ranked players are included in our sample. If our sample were extended to the top 180 players, the result may have been different.

Table 2. Summary of efficiency scores

	Player Efficiency		
	2022	2021	2020
N	52	52	52
Min	0.97	0.97	0.96
Max	1.00	1.00	1.00
Mean	0.99	0.99	0.99
Standard deviation	0.01	0.01	0.01

Table 3 summarizes the DEA bootstrapping result. Each player is re-sampled about 1,268 times in a 2,000 times bootstrapping and still is very efficient. The median score of each player is used for LGCM.

Table 3. Summary of efficiency bootstrapping result

Efficiency bootstrapping result			
Season	2022	2021	2020
Average # included in bootstrapping	1,268	1,269	1,268
Mean	0.9947	0.9931	0.9909
Median	0.9947	0.9928	0.9912

4.2 PCA

To make a variable that integrated two sub-factors for mental factor and technical factor, respectively, PCA was implemented. The result is illustrated in Table 4. Putt average is given a far greater weight than drive accuracy among mental factors, and the number of birdies also is given a greater weight than the number of eagles among technical factors. The percentages of variance explained are greater than 99%.

Table 4. PCA results for mental factor and technic factor

Season	Mental factor coefficient			Technical factor coefficient		
	Drive accuracy	Putt average	C.O.D.	Birdies	Eagles	C.O.D.
2022	0.0150	1.0000	0.9900	1.0000	0.0260	0.9970
2021	0.0080	1.0000	0.9930	1.0000	0.0260	0.9960
2020	0.0001	1.0000	0.9900	1.0000	0.0160	0.9990

Table 5 summarizes the calculated variables for mental factor and technical factor. It is possible to confirm a substantial variation among players. These variables are included in LGCM.

Table 5. Summary of PCA scores for mental factor and technic factor

Season	Factor	N	min	Max	mean	Standard deviation
2022	Mental	52	28.62	31.68	29.88	0.55
	Technical	52	101.00	396.23	263.49	53.82
2021	Mental	52	28.72	31.85	30.05	0.59
	Technical	52	96.03	343.03	249.12	49.86
2020	Mental	52	27.59	30.95	29.89	0.63
	Technical	52	32.00	303.06	144.30	47.88

4.3 LGCM

Because LGCM is based on a structural equation model, it is necessary to examine the reliability of efficiency factor, mental factor, technical factor, and money list factor. The Cronbach's αs of all are greater than 0.6, which implies that they are reliable. Moreover, an appropriate relationship between factors was confirmed through a correlation analysis.

LGCM was implemented in 8 different structural equation models: 1. Latent variables for factors, no mediator, intercept → intercept, intercept → slope, slope → intercept, and slope → slope; 2. Latent variables for factors, no mediator, intercept → intercept, and slope → slope; 3. Latent variables for factors, mental mediator, intercept → intercept, intercept → slope, slope → intercept, and slope → slope; 4. Latent variables for factors, mental mediator, intercept → intercept, and slope → slope (Preacher et al., 2008); 5. Latent variables for years, no mediator, intercept → intercept, intercept → slope, slope → intercept, and slope → slope; 6. Latent variables for years, no mediator, intercept → intercept, and slope → slope; 7. Latent variables for years, mental mediator, intercept → intercept, intercept → slope, slope → intercept, and slope → slope; 8. Latent variables for years, mental mediator, intercept → intercept, and slope → slope (Peterson et al., 2011).

Table 6 illustrates the LGCM result of models 2, 4, and 7 because they show us a better fit result and regression weight result than others. Most measures of fitting appear to be acceptable, though some such as NFI look slightly bad (Lee and Lim, 2015). Players' career affects the intercept of money list positively, the intercept and slope of technical affect those of money list positively, the intercept of mental affect those of money list negatively while the intercept of mental affect those of money list positively, which implies that players' inborn mental level may reduce the level of discretion and mental level obtained through playing golf affect the growth of money list. In model 7, the result that player intercept and slope affect money intercept and slope positively, respectively, represents that they include technical factors. The mediating effect of mental factors could not be confirmed.

Table 6. Result of LGCM

<table>
<tr><td colspan="7">2. Latent variables for factors, no mediator, intercept → intercept, and slope → slope</td></tr>
<tr><td>GFI</td><td>TLI</td><td>CFI</td><td>NFI</td><td>RMSEA</td><td>AGFI</td><td>χ^2/df</td></tr>
<tr><td>0.8</td><td>0.97</td><td>0.97</td><td>0.42</td><td>0.02</td><td>0.72</td><td rowspan="6">1.01</td></tr>
<tr><td colspan="3">Significant paths</td><td>B</td><td>S.E.</td><td>C.R.</td></tr>
<tr><td colspan="3">Money intercept ← Career</td><td>1.95*</td><td>1.16</td><td>1.68</td></tr>
<tr><td colspan="3">Money intercept ← Technical intercept</td><td>1.99***</td><td>0.72</td><td>2.76</td></tr>
<tr><td colspan="3">Money slope ← Technical slope</td><td>2.25**</td><td>1.10</td><td>2.05</td></tr>
<tr><td colspan="3">Money intercept ← Mental intercept</td><td>-26.66***</td><td>10.15</td><td>-2.63</td></tr>
<tr><td colspan="7">4. Latent variables for factors, mental mediator, intercept → intercept, and slope → slope</td></tr>
<tr><td>GFI</td><td>TLI</td><td>CFI</td><td>NFI</td><td>RMSEA</td><td>AGFI</td><td>χ^2/df</td></tr>
<tr><td>0.83</td><td>1.14</td><td>1</td><td>0.49</td><td>0</td><td>0.74</td><td rowspan="5">0.94</td></tr>
<tr><td colspan="3">Significant paths</td><td>B</td><td>S.E.</td><td>C.R.</td></tr>
<tr><td colspan="3">Money intercept ← Career</td><td>1.95*</td><td>1.17</td><td>1.67</td></tr>
<tr><td colspan="3">Money slope ← Technical slope</td><td>2.39**</td><td>1.11</td><td>2.16</td></tr>
<tr><td colspan="3">Money intercept ← Technical intercept</td><td>1.91***</td><td>0.65</td><td>2.95</td></tr>
<tr><td colspan="7">7. Latent variables for years, mental mediator, intercept → intercept, intercept → slope, slope → intercept, and slope → slope</td></tr>
<tr><td>GFI</td><td>TLI</td><td>CFI</td><td>NFI</td><td>RMSEA</td><td>AGFI</td><td>χ^2/df</td></tr>
<tr><td>0.8</td><td>0.93</td><td>0.95</td><td>0.42</td><td>0.03</td><td>0.72</td><td rowspan="6">1.03</td></tr>
<tr><td colspan="3">Significant paths</td><td>B</td><td>S.E.</td><td>C.R.</td></tr>
<tr><td colspan="3">Money intercept ← Player intercept</td><td>278.27***</td><td>64.89</td><td>4.29</td></tr>
<tr><td colspan="3">Money slope ← Player slope</td><td>534.66*</td><td>308.36</td><td>1.73</td></tr>
<tr><td colspan="3">Money slope ← Mental slope</td><td>161.32*</td><td>86.15</td><td>1.87</td></tr>
<tr><td colspan="3">Money intercept ← Mental intercept</td><td>-21.80*</td><td>12.79</td><td>-1.71</td></tr>
</table>

*: $p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$

Table 7 illustrates the mean of latent slope for each factor. The significant mean of latent slope represents that there would be a growth in players' factors. As seen in the table, Money list and technical factors are increasing over seasons, while efficiency is not changed, and mental slope shows very slight increases and decreases over seasons.

Table 7. Mean of slope of growth for factors

Money list slope		Efficiency slope		Mental slope		Technical slope	
$2 \rightarrow 3$	$1 \rightarrow 2$	$2 \rightarrow 3$	$1 \rightarrow 2$	$2 \rightarrow 3$	$1 \rightarrow 2$	$2 \rightarrow 3$	$1 \rightarrow 2$
1.10^{**}	1.73^{**}	0.00^{**}	0.00^{**}	-0.01^{**}	0.01^{**}	0.12^{**}	1.01^{**}

**: $p < 0.05$, 1: 2020 season, 2: 2021 season, 3: 2022 season.

5 Conclusion

This research aimed to investigate the differences between players ranked in the official money list, which lists the total award money won by LPGA players, and is their most important performance measure. To examine what factors affect it and whether they grows over seasons or not, an LGCM was developed and implemented using mental factor, efficiency factor, technical factor, and career of players.

We confirmed that the efficiency of the top-ranked players is very homogeneous, and in seasons and career, mental factors acquired by playing golf, technical factors affect placement on the money list positively but inborn mental factors affect it negatively. Based on these findings, this paper argues that players in the LPGA will naturally try to improve their techniques and mental power and accumulate the experience steadily to win more money.

In the future, this research should be extended by including more players in the analysis and performing a survey of players' psychological capital to measure mental factors accurately.

References

[1] Charnes, A., Cooper, W.W., Lewin, A.Y., and Seiford, L.M. (1994) *Data Envelopment Analysis: Theory, Methodology, and Applications*. Springer.

[2] Chung, S., Yeo, I., and Chung, Y. (2017) *The effect of technical and psychological factors on the performance and their official money for LPGA tour players*. Korea Wellness Journal **12**(4), 601-614.

[3] Chung, S. (2020) *Analysis of Structure Model for Performance in LPGA Players*. The Korea Journal of Sports Science **29**(5), 807-817.

[4] Fried, H.O., Lambrinos, J., and Tyner, J. (2004) *Evaluating the performance of professional golfers on the PGA, LPGA and SPGA tours*. European Journal of Operational Research **154**, 548–561.

[5] Kim, E. and Chae, J. (2021) *Comparison of the performance levels of US female professional golf players*. International Journal of Applied Sports Sciences **33**(1), 61-73.

[6] Kim, N. and Min, D. (2014) *The study of changes in performance in KLPGA using growth curve analysis*. Journal of Korean Data & Information Science Society **24**(5), 847-855.

[7] Lee, H. and Lim, J. (2015) *Structural equation model analysis and AMOS 22*. Jiphyunjae.

[8] Lind, D., Marchal, W., and Wathen, S. (2021) *Statistical techniques in business and economics*. McGraw-Hill.

[9] Min, N., Kwak, T., and Kim, E. (2022) *Comparison of Women's Professional Golf Performance in Korea and Asian Countries using LPGA Data*. International Journal of Coaching Science **24**(2), 262-271.

[10] Penick, H., Shark, B., and Love III, D. (2012) *Harvey Penick's little red book: Lessons and teachings from a lifetime in golf*. Simon & Schuster.
[11] Peterson, S.J., Luthans, F., Avolio, B.J., Walumbwa, F., and Zhang, Z. (2011) *Psychological capital and employee performance: A latent growth model approach*. Management Department Faculty Publications of University of Nebraska – Lincoln, Lincoln, NE.
[12] Preacher, K.J., Wichman, A.L., MacCallum, R.C., and Briggs, N.E. (2021) *Latent growth curve modeling*. Sage.

Scoring probability model based on service landing location and ranking points in men's professional tennis matches

Fumiya Shimizu* and Eiji Konaka*†

* 1-501, Shiogamaguchi, Tempaku-ku, Nagoya, Aichi, JAPAN. 468-8502.
† Corresponding author. Email address: konaka@meijo-u.ac.jp

Abstract

This study mainly aims to construct a mathematical model that expresses the scoring probability of service shots. The explanatory variables are the service landing location and the ranking points of the server and receiver (logarithms of ranking points for the two players). The produced model reveals the following facts: In the 1st service, advantageous locations for the servers are not symmetric between the ad and deuce sides. The higher-ranked player can win more points than lower-ranked players even if his shot lands in an easy location, that is, in the middle of the service court.

1 Introduction

Service is regarded as the most crucial shot in tennis matches because it is the only shot that can be played without being influenced by the opponent [1].

The following studies clearly and quantitatively show the services that can create advantageous situations and to what extent: Visualization of the boundary line of the advantageous/disadvantageous area of the service landing position based on the data of the service landing point and its corresponding scorer (i.e., server or receiver) [2]. A study that focuses only on the 1st serve of the Australian Open, training a prediction model that outputs whether it is a service ace by using a k-means method [3]. A study of the relationship between the landing position of the service and subsequent strokes and the scoring in women's professional tennis [4].

This study mainly aims to construct a mathematical model that takes the service landing location and the difference in strength as arguments and outputs the probability of scoring. The construction of a score probability prediction model that includes all of these factors has not been reported within the scope of the author's research.

The remainder of this paper is organized as follows. Chapter 2 describes the official rules and ranking point system in professional tennis, which is necessary for the discussion in this paper. Chapter 3 presents a detailed definition of the data and method used to construct a score probability prediction model. In Chapter 4, we construct a prediction model based on actual measurement data and then discuss the characteristics of each situation based on the output. Finally, Chapter 5 concludes the paper.

2 About Tennis

This chapter explains the rules and terminology of tennis, which are necessary for the discussion in this paper.

2.1 Official Rules and Basic Service Statistics

The names of the lines on a tennis court are presented in Figure 1.

In this study, the server and receiver are in the lower and upper parts of Figure 1, respectively. Therefore, the sevice ball flies from the bottom to the top in this figure. In this figure, the output of the prediction model is the boxes (service court) indicated by the bold line. On the service court, when the center mark is at the boundary, the right and left sides of the court are called the "deuce side" and the "advantage side," respectively. In this paper, the "rear" of the service area indicates the area near the baseline.

In the case of the first fault, the player may again serve from the same side, but if this is also a fault, the receiver scores a point (double fault).

Therefore, the 1st service is fast so as not to be returned well by the receiver because a second service can be awarded even if the first service is a fault. Conversely, the 2nd service is often a slower service that focuses on obtaining the ball into the service court because, if it does not go into the court, it will be a point for the opponent.

The velocity distribution was examined from the dataset of 104,100 services (67,342 and 36,758 1st and 2nd services, respectively) recorded at 41 Grand Slam tournaments [5]. The average speed of the 1st service was 173 [km/h], and the first and third quartiles were 158 and 189 [km/h], respectively. Conversely, the second service has the first to third quartiles of 131, 141, and 152 km/h. Therefore, there is a speed difference between the two services.

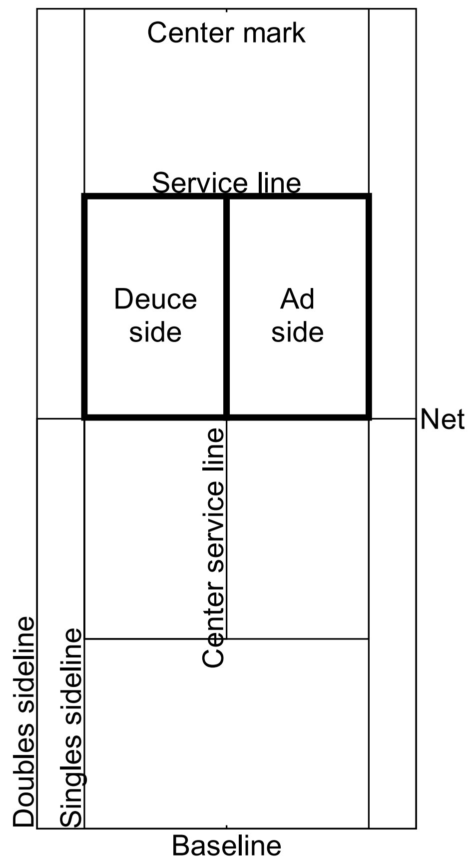


Figure 1: Tennis court and name of the lines

2.2 ATP Ranking System

ATP Tour players are awarded points based on their ranking in each tournament. The points awarded for each ranking in the tournaments are presented in the official website [6]. From the webpace, we can observe that the points awarded for each competition are multiplied by a factor of 5/3 to 2 for each victory.

Reference [7] demonstrates that the winning percentage in the ATP Tour from 2009 to 2015 fits well with a logistic regression model with the log of the ranking point ratio as an input variable. Reference [8] compares the predictive performance of official rankings with several statistical methods using win/lose and scoring information. The results show that these evaluation methods are comparable and that the official ranking is a superior performance indicator.

Based on these findings, it is appropriate to use the logarithm of the ranking point ratio as an indicator of the strength difference between the server and receiver.

3 Service Scoring Probability Modeling

This chapter describes the data and methods used in the model construction. A list of variables used in this paper is presented in Table 1.

3.1 Data set

The model is built using the data from 897 matches from 30 of the 62 official ATP Tour tournaments in 2019, and data on service landing locations are available in the official website [9].

Table 1: List of variables

Variable	Definition
s,r	Index of server and receiver
x,y	Searvice landing point $y \in [450,692]$, $x \in [50,208]$ (ad side), $x \in [206,364]$ (deuce side)
P_s,P_r	Ranking point
$L_{s,r}$	Logarithm of ranking point ratio $\log(P_s/P_r)$
t	Score. Server:1, Receiver:0
$(x,y,L_{s,r},t)$	Data decord of one service
$w(x,y,L_{s,r})$	Scoring probability
$\hat{w}(x,y,L_{s,r})$	Predicted scoring probability

From these public data, we obtain the landing location (x,y) of each service that entered the service court, its score $t \in \{0,1\}$, and server and receiver names. The ranking points of the server and receiver at the time of the competition are retrieved from the formatted dataset [5].

3.2 Characteristics of Service Distribution and Data Preprocessing

Figure 2 presents the actual distribution of the service landing locations.

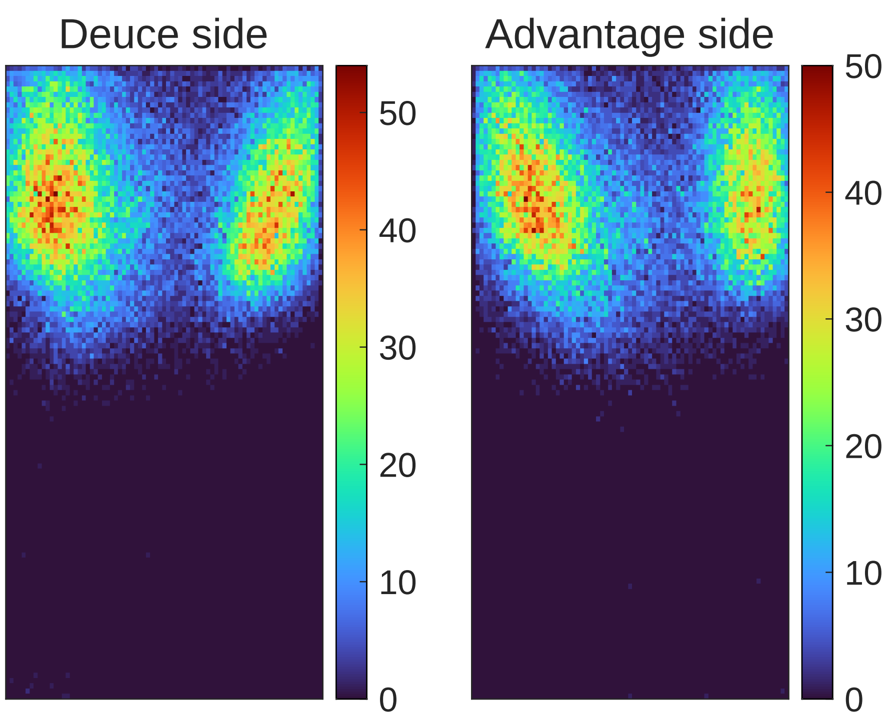


Figure 2: Distribution of service landing locations

The total number of services is 111,473 (59,188 and 52,285 for deuce and advantage sides, respectively). The distribution is different between the deuce and advantage sides (not symmetrical either). Therefore, a predictive scoring probability model should be constructed for each side.

Few services landed on the service court near the net. Specifically, on both sides, less than 1% services landed in 1/2 of the service court near the net.

3.3 Scoring Probability Model Construction Method

The following is the algorithm for constructing the predicted scoring probability for the landing point of the service. This algorithm is based on a stepwise generalized linear regression implemented in MATLAB [10].

1. Using service-wise data [5][9], construct a database with each service as a record (x, y, P_s, P_r, t) (service landing location, ranking points, and scores).

2. Dummy data with $t = 0$ are added equally to the 1/5 of the service court near the net. At this time, we assume that no difference exists in the players' strength and set the logarithm of the ranking point ratio $L_{s,r} = \log(P_s/P_r) = 0$.

3. Construct a generalized linear regression model with $(x, y, L_{s,r})$ as the predictor variable, t as the response variable, and the starting model as a polynomial [10]. The maximum order of the terms included in the prediction model was set four, the distribution of the response variable was set to binomial distribution, and the other settings were a default value of `stepwiseglm`.

Using this method, the 1st/2nd services, as well as the deuce and advantage sides, are split to build four models.

4 Model Construction Result

The model construction resulted in a logistic regression model containing terms of the highest fourth order for each of $(x, y, L_{s,r})$. The details of the model are omitted due to space limitation.

The predicted scoring probabilities for the 1st services on each side are presented in Figure 3. In this case, the ranking point ratio is assumed $P_s/P_r = 1$, where there is no difference in players' strength.

In Figure 3, the areas of the high scoring probability are spread around the sideline and center line. This figure presents the similar characteristics on both sides in the rear 1/3 of the service court. Specifically, the deuce side has the $\hat{w} = 0.65$ line protruding from the left (outside) toward the center, whereas the advantage side has the line with the same value protruding from the right (outside) toward the center. Both of these advantageous areas are from the sideline. Therefore, aiming at the sideline for the 1st service, the player has a better chance of advancing in the game than by aiming at the center line side, even if the ball slightly lands to the inside. Simultaneously, the probabilities have asymmetric nature. For instance, the shape and size of $\hat{w} \geq 0.65$ is different on both sides. On the advantage side, this area protrudes to the center of the service area, whereas smaller protruded area exists on the deuce side. We suppose that the reason for this phenomenon is biased dominant arms of the servers. This issue is our fiture work.

We can observe symmetrical similarities in the probability of scoring near the service line. On both sides, the probability of scoring near the center line increased as the landing location moved closer to the service line. The possible reason is that the further the service lands toward the center line, the more the receiver is forced far from the service line. Therefore, the receiver has a difficulty returning the angled return. Additionaly, by forcing the receiver far from the service line, net play would be a possible choice for the server.

Near the sideline, we can observe the maximum probability of scoring at a location where the service landing location moves 1/4 length of the service court from the service line. The service landed in this area can move the receiver far from the sideline. Therefore, return for the reverse side would be a good choice for the server.

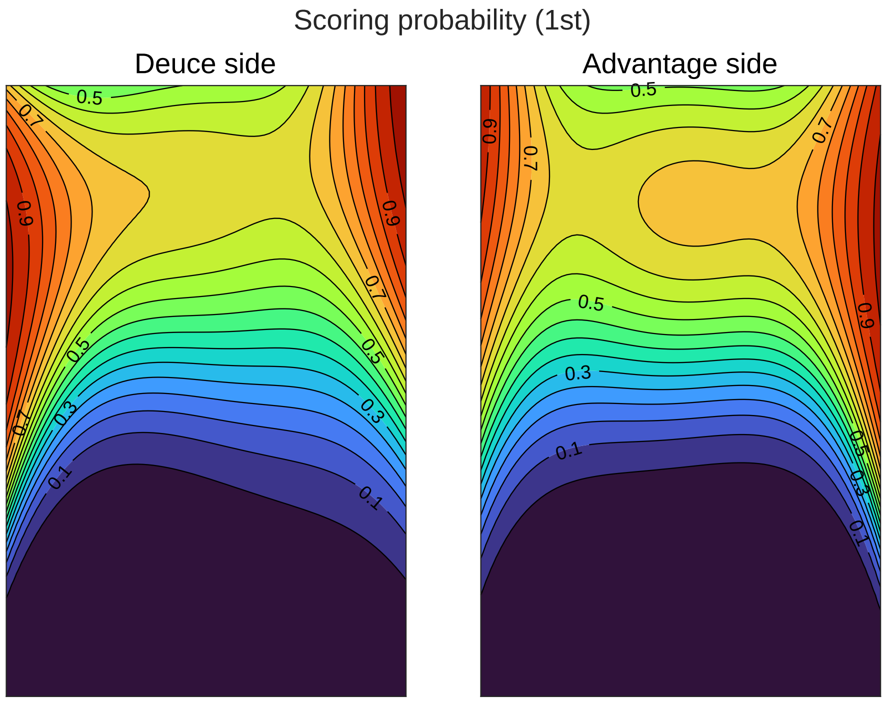


Figure 3: Scoring probability of 1st service

5 Conclusion

In this study, a stepwise regression method was used to construct a scoring probability prediction model based on service landing and ranking points in men's professional tennis. Data from 111,473 services in 897 matches were used to construct the model. The result indicates that visually clarifying the intuitive knowledge of "what kind of service produces what kind of advantageous situation" is possible by supporting it with mathematical evidence. The heat maps created using this model were compared by situation. The results clearly indicated the areas of advantage for each service and side.

References

[1] C. Johnson and M. McHugh. Performance demands of professional male tennis player. *Br J Sports Med 40*, page 696?699, 08 2006.

[2] `https://logmi.jp/tech/articles/324033`. (Written in Japanese. Accessed on 01/14/2022).

[3] David Whiteside and Machar Reid. Spatial characteristics of professional tennis serves with implications for serving aces: A machine learning approach. *Journal of Sports Sciences*, 35(7):648–654, 2017. PMID: 27189847.

[4] Philipp Born, Louis Malejka, Marius Behrens, Ralph Grambow, Dominik Meffert, Jonas Breuer, and Tobias Vogt. Stroke placement in women's professional tennis: What's after the serve? *Sport Science*, 3:37–45, 08 2021.

[5] Jeff Sackmann. JeffSackmann. GitHub. `https://github.com/JeffSackmann`. (Accessed on 10/28/2021).

[6] Rankings | FAQ | ATP Tour | Tennis. `https://www.atptour.com/en/rankings/rankings-faq`. (Accessed on 01/15/2022).

[7] Eiji Konaka. Match results prediction ability of official ATP singles ranking. preprint on arXiv, https://arxiv.org/abs/1705.05831, May 2017.

[8] Stephanie Ann Kovalchik. Searching for the GOAT of tennis win prediction. *Journal of Quantitative Analysis in Sports*, 12(3):127–138, 2016.

[9] Results Archive | ATP Tour | Tennis. `https://www.atptour.com/en/scores/results-archive?year=2019`. (Accessed on 10/28/2021).

[10] Mathworks. stepwiseglm. Create generalized linear regression model by stepwise regression. `https://jp.mathworks.com/help/stats/stepwiseglm.html#d123e768995`. (Accessed on 3/23/2023).

Pitcher selection scenarios in baseball: A multi-criteria approach

Marina V. Polyashuk

Northeastern Illinois University, Chicago, IL 60625, USA, email m-polyashuk@neiu.edu

Abstract

This study focuses on various scenarios that are typical for a Major League Baseball (MLB) manager's routine decisions in pitcher selection. In various types of scenarios, we combine practical reasoning and available statistical data with the application of multi-criteria decision analysis to create mathematical models. The suggested models and methodologies are aimed at informing, supporting, and simplifying baseball manager's decisions in pitcher selections. Throughout the paper, the proposed methodologies are applied to the scenarios from the 2022 season of the Chicago White Sox team, and actual managerial decisions are compared to the solutions proposed by our models.

1 Introduction and a brief review of existing research

There are various decisions that a Major League Baseball (MLB) team manager makes prior to and during every of the 162 games of the season. This paper focuses on pitcher selection decisions. In any of the scenarios when a pitcher is selected, managerial decisions may be made based on the institutional knowledge of the game of baseball, personal experience in similar cases, statistical analysis (such as sabermetrics), or all the above.

Our view is that allowing managers to deal with a narrower set of alternative decisions by excluding non-optimal choices is a valuable tool in informing and supporting managerial decisions. In this paper, we will offer models that would not replace the manager in the decision-making process; rather, our approach is to use mathematical models in order to narrow the number of viable alternative decisions, which would support baseball managers' decision process and make them part of it. We will do this by formalizing several types of pitcher selection situations by means of multi-criteria decision analysis.

With regard to the theoretical publications aimed at the problems of pitcher selection in baseball, some of the most prominent examples create detailed mathematical models of an entire baseball game as Markov-chain models based on an earlier paper (D'Esopo & Lefkowitz, 1960). These models are used to evaluate pitcher management strategies, consider various aspects of pitcher substitution strategies and their effectiveness (see Bukiet 2018, Bukiet et al. 1997, Hirotsu & Wright 2004a; Hirotsu & Wright 2004b; Sidhu & Caffo, 2014). While simplifying some aspects of the baseball game outcomes, the models introduced in the mentioned research are complex and difficult to apply as a routine procedure by baseball managers in real time. In some cases, the random processes used to model a baseball game have hundreds of states and thousands of transition probabilities in their transition matrices; it is hardly possible for a baseball manager to fully understand such models and be part of the decision making process. In other notable publications, the major goal is to decide if it is beneficial to replace the current pitcher by a reliever (Soto-Valeroa et al. (2017)). In Harrison & Salmon (2019), analytics and clustering analysis is used effectively to determine better pitcher-batter matchups. Academic statistical studies in sports (Albert & Koning, 2007; Albert & Bennett, 2003) offer various interpretations of statistical data related to pitchers' performance.

Among many ways of evaluating pitcher performance, there are traditional statistics as well as newer, comprehensive measurements designed by Sabermetrics followers to be meaningful predictors of pitcher success or failure. Among the older measures evaluating pitcher performance is *ERA*, the earned run average, the average number of runs allowed by the pitcher in a hypothetical 9-inning game. However, the problem with this and other traditional measures is that they are strongly influenced by the team's defense and do not represent pitcher's unique contribution, given that the team's players greatly contribute to the *ERA*'s either lower (better defense) or higher (worse defense) values. Over the years, sabermetrics authors introduced several composite characteristics of pitcher performance that, arguably, represent better evaluations of the game outcomes caused by pitcher alone. One such measure is *FIP* (fielding independent pitching), which falls in about the same range as a typical *ERA*. *FIP*, the measure invented by Tom Tango in 2008, is considered as one of the best ways to evaluate overall pitcher performance.

2 Rotational selection scenarios and multi-criteria models for decision support

Rotational selection is a selection of the order in which starting pitchers are assigned in consecutive games. There are several situations during the baseball season when a rotational selection has to be made. Let us go over these possibilities.

1. Before the season begins (or before the All-Star break), the starting rotation is decided. Therefore, a sequence of five starting pitchers is chosen, with every pitcher working their game after at least four days of rest. Traditionally, this sequence is set in a descending order of pitchers' strength, ability, and talent, with "aces" being put in the lead of the pitching rotation.
2. Any time during the season when the team has a day off, manager may change the order in pitcher rotation without forcing any pitcher to forgo his four days of rest.
3. On a day of a double-header (when the team plays two games on the same day), the manager has to choose an extra starting pitcher for one of the two games. It could be one of the relievers who are currently on the team roster or a pitcher from a Minor League team in the franchise who is called specifically for the given game and then returned to his team.

At this point, we will assume that the set of five starting pitchers is determined, and the manager's decision in this scenario concerns the sequence of the rotation of the five starters in a descending order of pitchers' strength, versatility, and talent. In general, the rotation can be established in $5!=120$ possible ways, however using a conventional approach, we can instead consider the decision space consisting of the given five starting pitchers whom we wish to rank. To set up a criteria space, we consider what makes a good starting pitcher based on a concept of a "quality start", which is defined as pitching at least 6 innings and allowing no more than 3 earned runs. This concept allows us to point out two major criteria for pitcher evaluation: the average number of runs allowed by the pitcher in a hypothetical complete game (best represented by *FIP*), and the average number of innings pitched in a start (*IPS*). In a bi-criteria model with these two criteria forming the criteria space, we can define a preference relation that will help to rank the pitchers.

Definition 1

Given the set of pitchers $P \subseteq D$, where D is the decision space consisting of all pitchers on the team roster and $P = \{p_1, p_2, p_3, p_4, p_5\}$ is the given set of starters, let $F_1(p_i) = FIP(p_i)$ and $F_2(p_i) = IPS(p_i)$,

Then the problem of pitcher selection in Rotational Scenario A is to: $\begin{cases} \min\limits_{p_i \in P} F_1(p_i) \\ \max\limits_{p_i \in P} F_2(p_i) \end{cases}$

This problem formulation suggests that the goal in the decision making process is to simultaneously minimize the number of runs allowed by the pitcher and to maximize the number of innings that the pitcher works per start. Obviously, it is not always possible to find a single solution that satisfies both goals; however, it usually helps to eliminate the less perfect choices and facilitate the decision making process.

Let us consider an example: the starting pitcher rotation for the Chicago White Sox team as it stood at the beginning of the 2022 season. In the figure shown below, the criteria space based on the formulated above bi-criteria problem contains five points whose coordinates represent *FIP* and *IPS* for the five 2022 Chicago starting pitchers Dylan Cease, Lucas Giolito, Dallas Keuchel, Michael Kopech, and Vince Velasquez, based on their performance in 2021. If the manager took a quick look at Figure 1, it would be easy to visualize the preferred three non-dominated points located closer to the left top corner of the graph: $(3.45,\ 5.2)$ (Dylan Cease), $(3.84,\ 5.8)$ (Lucas Giolito), and $(3.01,\ 3.5)$ (Michael Kopech).

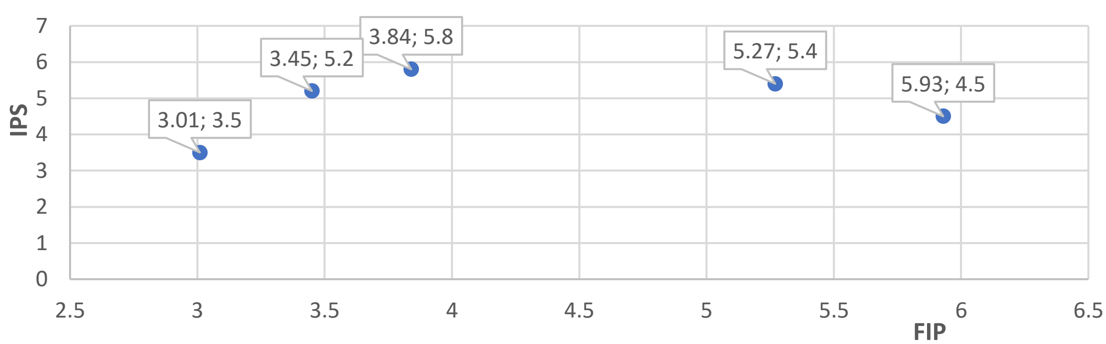


Figure 1: Evaluations for five starting pitchers for the 2022 Chicago White Sox

Based on the simple instrument demonstrated above, the number of feasible starting rotations is reduced from 120 to just 6 optimal choices. These options are generated by permuting the set of three non-dominated alternatives while keeping the other two alternatives in descending order of preference. Notably, at the beginning of the 2022 season the pitcher rotation looked almost exactly like this.

3 Relief selection scenarios

When the starting pitcher has expended a large number of pitches (usually, about 100) or begins to falter and becomes inefficient in taking out the batters of the opposing team, then the manager selects a replacement from the bull-pen. This can happen in several different ways, where each case is defined by the number of outs and the number and locations of the runners of the opposing team ("base-out" situations). In the table below, we can observe all possible cases; we have also placed the expected number of runs in the remaining part of the inning for each case (using the Run Expectancy Matrix provided in Albert, 2020, which is based on the large sample of statistical data from the entire 2019 MLB season).

0 outs, 0 runners **0.53**	0 outs, 1 runner on base 1 **0.94**	0 outs, 1 runner on base 2 **1.17**	0 outs, 1 runner on base 3 **1.43**	0 outs, 2 runners on bases 1 and 2 **1.55**	0 outs, 2 runners on bases 1 and 3 **1.80**	0 outs, 2 runners on bases 2 and 3 **2.04**	0 outs, 3 runners on bases 1, 2, and 3 **2.32**
1 out, 0 runners **0.29**	1 out, 1 runner on base 1 **0.56**	1 out, 1 runner on base 2 **0.72**	1 out, 1 runner on base 3 **1.00**	1 out, 2 runners on bases 1 and 2 **1.00**	1 out, 2 runners on bases 1 and 3 **1.23**	1 out, 2 runners on bases 2 and 3 **1.42**	1 out, 3 runners on bases 1, 2, and 3 **1.63**
2 outs, 0 runners **0.11**	2 outs, 1 runner on base 1 **0.24**	2 outs, 1 runner on base 2 **0.33**	2 outs, 1 runner on base 3 **0.38**	2 outs, 2 runners on bases 1 and 2 **0.46**	2 outs, 2 runners on bases 1 and 3 **0.54**	2 outs, 2 runners on bases 2 and 3 **0.60**	2 outs, 3 runners on bases 1, 2, and 3 **0.77**

Table 1: Base-Out Situations with Run Expectancies

It is possible to identify the following situations in which a reliever is called.

1. A reliever is called at the start of an inning, when the bases are empty. This is a situation (1 of 24 possibilities) when the current pitcher has finished an inning of work and is being replaced.
2. A reliever is called in the middle of an inning (so it is not the first scenario) when there are 2 outs or no runners, or there is 1 out and 1 runner on base 1 or base 2. This scenario combines 11 of 24 possibilities (the corresponding cells in the table are shaded lighter grey).
3. A reliever is called in the middle of an inning when there are no outs and at least one runner on any base or there is 1 out and at least one runner in scoring position (base 2 or 3); this scenario accounts for 12 of 24 possibilities (the corresponding cells in the table are shaded darker grey).

Based on the expected number of runs scored in the remaining part of the inning, the first and the second scenarios may be considered similar, with the expected number of runs considerably below 1 (with the average of 0.46), while the third scenario is prone to produce a higher number of runs (with the average of 1.46) for the opposing team, is much tougher for the pitcher and requires different pitching strategies. To develop mathematical models that would allow us to effectively select the most suitable pitcher for each of these scenarios, we will define the combination of cases (1) and (2) as Relief Scenario A, and case (3) as Relief Scenario B**.** For the first type of scenario, the most valued qualities for such a reliever are similar to those for a starting pitcher. However, *IPS* (innings per start) criterion must be replaced by another measure of pitcher's effectiveness; *P/IP* (the number of pitches per inning pitched) is an example of such measure.

Definition 2

Given the set of pitchers $P \subseteq D$, where D is the decision space consisting of all pitchers on the team roster and $P = \{p_1, p_2, \ldots, p_k\}$ is the given set of available relievers on a given day, let $F_1(p_i) = FIP(p_i)$ and $F_2(p_i) = P/IP(p_i)$. Then the problem of pitcher selection in Relief Scenario A is to: $\begin{cases} \min\limits_{p_i \in P} F_1(p_i) \\ \min\limits_{p_i \in P} F_2(p_i) \end{cases}$.

It is clear from this definition that optimizing this type of pitcher selection can be modeled as a bicriteria problem similar to the one formulated in section 2, except that both criteria values have to be minimized.

The chart below shows an example from the Chicago White Sox 2022 season to demonstrate such scenario and possible optimal selections. The best option for Scenario A in 2022 was Reynaldo Lopez who had the minimum values with respect to both criteria, with 1.94 *FIP* value and 15.17 *P/IP* value. Next in the ranking induced by the above preference relation were Liam Hendricks (*FIP* of 2.75 and *P/IP* of 16.04) followed by Kendall Graveman (3.38, 15.9).

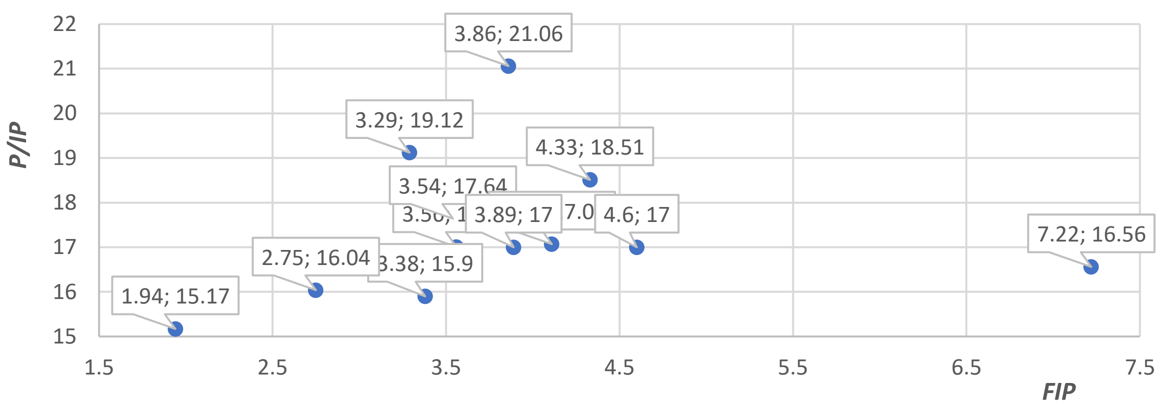


Figure 2: Relief pitching for the 2022 Chicago White Sox

The second scenario calls for a pitcher who could help in a difficult situation and prevent the opposing team's runners to score. Among the runs scored, it is reasonable to distinguish the percentage of inherited runners that scored a run (*IS%*) as the second criterion evaluating pitchers' effectiveness. In this case, we can consider the bi-criteria model (similar to the one given in Definition 2) where we try to simultaneously minimize pitcher's *FIP* and *IS%*.

In the example of the Chicago White Sox 2022 season, we can discern the following optimal selections for Scenario B: Reynaldo Lopez who had the minimum value with respect to *FIP* (1.94) and Kendall Graveman who had the minimum *IS%* (25%). It turns out that Reynaldo Lopez played most innings in relief in 2022, which actually reflects his high value in short, as well as in long relief outings during that season.

4 Summary and conclusions

The goal of this work is to provide a research-based, yet simple and non-intrusive, way to support a baseball manager or analyst making pitcher selection decisions. This can be achieved through a process of narrowing of the set of alternative decisions by excluding non-optimal choices. Conveniently, bi-criteria models with preference binary relations do just that.

In this paper, we focused on pitcher selection in two of types of scenarios: rotational and relief pitcher selections. Further, we developed bi-criteria models, each with a binary preference relation; our choice of such models has been motivated by a desire to make them easily understandable on both analytical and visual levels. In many ways, our model results were consistent with the actual managerial decisions made in the course of the 2022 season of the Chicago White Sox.

Ultimately, we argue that the use of the methodology presented here in addition to the direct insight that baseball managers have about their players' specific conditions, capabilities, and availabilities is the best approach to finding optimal solutions. Going forward, it would be useful to discern how accurate and helpful our models are by consistently applying them in an experimental one-season study in cooperation with a team management.

References

[1] J. Albert (2020) Summarizing a Run Expectancy Matrix *BaseballwithR.wordpress.com* (online).

[2] J. Albert and R. D. Koning (eds) Statistical Thinking in Sports (Chapman and Hall/CRC: New York)

[3] J. Albert and J. Bennett (2003) Curve Ball: Baseball, Statistics, and the Role of Chance in the Game, revised edition (Copernicus Books: New York).

[4] B. Bukiet (2018) Mathematically Modeling Baseball *Notices of the AMS*, **65**, 428-430.

[5] B. Bukiet, E. R. Harold, and J. L. Palacios (1997) A Markov Chain Approach to Baseball *Operations Research*, **45**, 14-23.

[6] D. A. D'Esopo and B. Lefkowitz (1960) The Distribution of Runs in the Game of Baseball *SRI Internal Report*.

[7] W. K. Harrison and J. L. Salmon (2019) Leveraging pitcher-batter matchups for optimal game strategy *Proc. MIT Sloan Sports Analytics Conference*, 1–22.

[8] N. Hirotsu and M. Wright (2004) Modeling a Baseball Game to Optimize Pitcher Substitution Strategies Using Dynamic Programming, in: Butenko, S., Gil-Lafuente, J., Pardalos, P.M. (eds.) *Economics, Management and Optimization in Sports*. Springer, Berlin, Heidelberg.

[9] N. Hirotsu and M. Wright (2004) Modeling a Baseball Game to Optimize Pitcher Substitution Strategies Incorporating Handedness of Players *IMA Journal of Management Mathematics*, **16**, 179-194.

[10] G. Sidhu, and B. Caffo (2014) "MONEYBaRL: Exploiting pitcher decision-making using reinforcement learning," in *The Annals of Applied Statistics*, **8**, 926–955.

[11] C. Soto-Valeroa, M. González-Castellanosa and I. Pérez-Morales (2017) A predictive model for analysing the starting pitchers' performance using time series classification methods *International Journal of Performance Analysis in Sport*, **17**, 1-18.

A novel method for player selection in T20 cricket

Bhaskar Basu * and Jimut Bahan Chakrabarty**

* Xavier Institute of Management, XIM University, Bhubaneswar (India), email address: bhaskar@xim.edu.in
** Indian Institute of Management, Sambalpur, email address: jimutc@iimsambalpur.ac.in

Abstract

Cricket as a sport has evolved over time and newer formats have emerged considering the changing tastes of the aficionados of the game. T20 cricket is the most recent and most popular international format of cricket. The fast-paced nature of the format demands teams to play with a different strategy and has scope for innovation compared to the longer formats of the sport. The commercial viability and brand value of the Indian Premiere League (IPL) has made the format popular among both players and spectators over the years. Under such circumstances, the traditional metrics to assess players' capabilities requires a fresh look. This paper aims to develop a novel method for player selection in T20 cricket by assessing players' performances in particular phases of the game using a proposed mathematical model. Adopting a dynamic approach of generating scores through the proposed model on a match-to-match basis, this article will help determine the best-equipped players on the field. The proposed model will be tested using most recent data from IPLT20 and international T20 cricket series played by India.

1 Introduction

Cricket is an outdoor spectator sport played on a 22-yard hard surface of earth called the pitch in an oval ground surrounded by spectator stands and players dressing room and other amenities for the spectators. The game is primarily between two teams of 11 players each vying for supremacy with the bat and ball governed by certain rules and regulations. The game of cricket has a long history dating back to the 16th century, but T20 cricket as a format gained popularity in the English domestic league around 2003 (Davis et al., 2015). Till now, eight T20 World Cups having been contested (2007, 2009, 2010, 2012, 2014, 2016, 2021, 2022) among international teams playing cricket. The popularity of T20 cricket exploded in India after it won the inaugural World Cup in South Africa and the advent of the Indian Premiere League (IPL) initiated by the Board of Control for Cricket in India (BCCI) in 2008.

A balanced cricket team is a blend of specialist batsmen, all-rounders (who can both bat and bowl), a wicketkeeper and specialist bowlers, led by a captain amongst them. Barring subtle differences like fielding restrictions and limits on the number of overs bowled by a bowler, T20 cricket follows most of the rules followed in one-day cricket. The key difference is the total number of overs bowled in a game (50 overs vs 20 overs), but the shorter span of time of a T20 cricket match attracts lot more spectators to the game, translating into higher revenues for the sports club or federation. This has led to the increased scheduling of matches in this format both in domestic competitions and international matches between cricket playing countries. The prize money of the winning team and individual awards for performers in these matches are also attractive. There is intense rivalry among teams to bid for players in IPL auctions arranged periodically by BCCI. The right selection of individual players (batsmen, all-rounders, wicketkeepers and bowlers) to make a winning team is one of the biggest challenges of a franchisee or team management. Good, consistent performance of the cricket players can turn them into stars pursued by fans and brands alike, leading to alternate revenue stream like endorsements. The coaches, management and captain of the franchisee have the onus to select the best playing eleven from the squad depending on the opposition and the roles provided to the players. As such, the cumulative performance statistics of the individual players becomes very critical in the team combination. Some of the concerns regarding selection of the final playing eleven includes injury concerns or fitness related issues, lack of form or confidence and familiarity or unfamiliarity with the opposition.

2 Literature review

“During the past decade a large number of academic papers have been published on cricket performance measures and predictive methods” (Lemmer, 2011a). Any cricket match generates vast amount of data and so there are several quantitative research works on cricket across different formats. Lately there have been several predictive models proposed for the winning team of a league or tournament. Selected quantitative works have been examined for the methodology adopted and summarized in the following section.

Kamble et al. (2011) used AHP to select players constituting a team based on their unique skill of batting, bowling or fielding. Ahmed et al. (2012) used evolutionary multi-objective optimization to choose players constituting a cricket team. Amin and Sharma (2014) proposed a two-stage method for measuring and ranking batting parameters in T20 cricket using ordered weighted averaging (OWA) operator and regression. The important parameters chosen by the researchers for the batsmen included Highest Score (HS) in a match, Average (Avg), Strike Rate (S/R), numbers of 4s and 6s hit by a batsman. Lemmer (2002) carried out similar studies with bowlers, factoring the average number of runs conceded per wicket taken (A), the economy rate (E), which is the average number of runs conceded per over bowled, and the strike rate (S), which is the average number of balls bowled per wicket taken. Lemmer has also contributed to measuring the fielding performance in a match (Saikia et al. 2012; Saikia et al., 2017) and evaluating the performance of wicketkeepers (Lemmer, 2011b).

Davis et al. (2015) developed a T20 simulator that calculated the probability of a first-innings batting outcomes dependent on batsmen, bowler, and number of overs consumed and total wickets lost. Brown, Saikia et al. (2017) evaluated the performance of the players into a single numerical value, which was a measure of the cricketer’s cricketing efficiency. Sharma (2013) investigated the systematic covariation among various dimensions pertaining to batting and bowling capabilities of T20 cricket using factor analysis and concluded that batting capability dominates over bowling capability. Saikia et al. (2019) used the binary integer programming for selecting the optimum balanced cricket team from a host of cricketers of varied expertise. Swartz et al. (2009) used the Bayesian log-linear model to develop the most optimal batting order in one-day cricket. Fuzzy logic and stochastic models were applied to model the uncertainties involved in the measure of performance of cricketers (Damodaran,2006; Singh et al., 2011).

Principal component Analysis (PCA) is a wide area in statistical science and has been used with cricket data in the past (Premkumar, Chakrabarty and Chowdhury, 2020; Prakash and Verma, 2022; Das et al., 2023). Lately, Gupta (2022) used PCA to evaluate batting skills of women cricketers’ performances in one-day internationals. Most researchers have considered T20 as a subset of the one-day 50 over format rather than a specialized stand-alone format for a balanced cricket team selection. In the model proposed by us, we try to retain traditional metrics, but apply it to specific phases of the match with varying weightages to maximize the output for the team with respect to the cricketer’s role in that phase.

3 Strategy building

Modern T20 has evolved over the years and there is a lot riding for a cricketer in the short time span of his playing at the highest level. He needs to stabilize his career with the franchisee who pays good money expecting him to deliver consistently for the team. The other motivation for any cricketer is cementing his place in the national squad through consistent performances, leading to enhanced reputation and possibly additional monetary benefits through brand endorsements. With so much money and popularity riding in the game, some players have made themselves unavailable to represent their country in order to play T20 leagues around the world.

The novelty proposed in this paper is to look at a modern T20 game as a game of three (3) phases, where each cricketer has been given a specific role to perform. The 3 phases are broken down as follows:

Phase 1-Overs 1-6 (Power Play)

Phase 2-Middle overs 7-15

Phase 3- Overs 16-20

The players have to be assigned specific roles based on their past performances during these phases, rather than looking at their performances in totality. For example, Jos Butler has scored big in the opening slot both for his country (England) and his franchisee, Rajasthan Royals. So, he should ideally revel in the opener's role and provide a brisk start in Phase-1. Similarly, Suryakant Yadav has been very consistent coming at number 4 in the batting order for India and his franchisee, Mumbai Indians. He gives much needed momentum to the team in the middle overs corresponding to Phase-2. Similarly, we can bank on big hitting all-rounders like Cameroon Green (Australia and Mumbai Indians) to do the job in the death overs corresponding to Phase-3. On the other hand, opening bowlers like Md Shami (India and Gujarat Titans) get movement early on with the new ball and have the ability to consistently pick wickets in Phase-1 and teams should utilize the bowling quota to the maximum in this phase. Spinners like Rashid Khan (Afghanistan and Gujarat Titans) have the ability to take wickets during the middle phase (Phase-2) and at the same time have a good economy rate to put pressure on opposing team. The last phase (Phase-3) is very challenging for the bowling team and they should resort to death specialist bowlers like Haris Rauf (Pakistan and Melbourne Stars) who have the ability to restrict runs and take wickets consistently in this phase.

4 Methodology

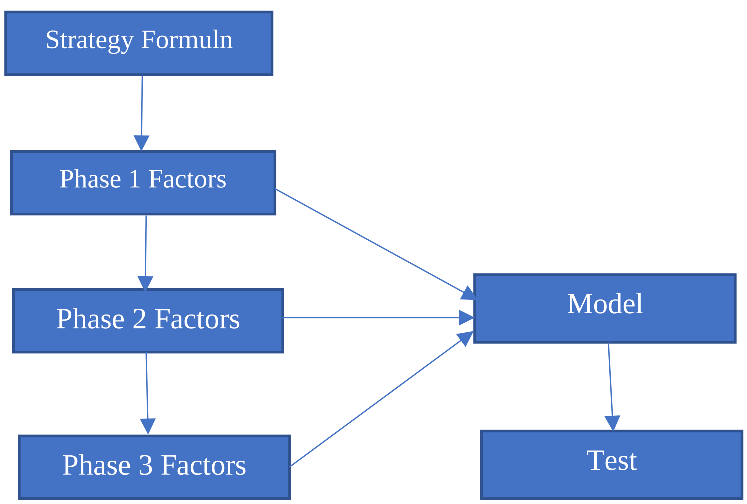


Figure 1: Novel Method for Player selection

The methodology to be followed for building the mathematical model and subsequent testing on actual data is depicted in Figure 1. The essence of the methodology adopted is that the performance of the cricket players will be evaluated in the role that they are most accustomed while playing for their respective franchisee/club/country. So, while evaluating the performance of a cricketer, we will select the best player for that particular position in our squad, assuming all the players are available for selection.

5 Model framework

The focus of this paper is to develop an index that is suitable to rank players based on their performance at different phases of a T20 game and thereby identify the best playing 11 for that team. Typically, a T20 game can be divided into three phases - Power play (1-6), Middle overs (7-15), and Death overs (16-20). Since the competencies of players performing in these three phases are different, one needs to ensure that the phases are considered while evaluating performance. Therefore, this paper identifies different variables for batsmen and bowlers to zero down on the best team composition based on the three phases of the game.

The variables identified for batsmen for different phases of the game are given below:

1. *Relative performance indicator*: Runs scored by a batsman on a match-by-match basis divided by all the runs scored by all the batsmen who have played in the tournament during the respective phase.
2. *Venue score*: Runs scored by a batsman in a particular match divided by the runs scored by all the batsmen in that particular venue during the respective phase of the game.
3. *Missing rate indicator:* Ratio of the number of balls missed by both teams in a match during a particular phase to the number of balls missed by the batsman under consideration.
4. *Scoring rate indicator:* Strike rate multiplied by the number of run scoring balls in a particular phase.
5. *Opposition Impact:* Ratio of the number of boundaries (fours plus sixes) scored by a batsman divided by the number of dot balls faced to the total number of boundaries (fours plus sixes) scored by the opposition by the number of dot balls faced during a particular phase of the match.
6. *Impact of Innings:* Runs scored by a batsman in an innings in a particular phase divided by runs scored by all the batsmen in that innings in that particular phase.

The variables identified for bowlers for different phases of the game are given below:

1. *Relative performance indicator:* Wickets taken by a bowler on a match-by-match basis divided by all the wickets taken by all the bowlers who have played in the tournament during the respective phase.
2. *Venue score*: Runs conceded by a bowler in a particular match divided by the runs conceded by all the bowlers in that particular venue during the respective phase of the game.
3. *Dot Ball indicator*: Ratio of total number of dot balls bowled by both teams in a match during a particular phase to the number of dot balls by the bowler under consideration.
4. *Strike rate impact:* Ratio of number of wickets taken+1/number of bowls bowled in that phase to the total no. of wickets+1 fallen against bowls delivered during that phase
5. *No ball impact:* Number of runs conceded off no- ball and the next delivery (free-hit) * (wicket +1)
6. *Opposition impact:* Ratio of the number of boundaries (fours plus sixes) conceded by a bowler divided by the number of dot balls bowled by him and the total number of boundaries (fours plus sixes) scored by the opposition to the number of dot balls faced during a particular phase of the match.
7. *Innings impact:* Economy rate by a bowler in a particular phase of an innings divided by economy rate of all the bowlers in that particular phase of the innings.

We use the factor analysis method to develop an index using the aforementioned variables. To describe it mathematically, let us consider an observable random vector $\boldsymbol{Z}$ with q components. Suppose the mean vector is given by $\boldsymbol{M}$ and $\boldsymbol{\Delta}$ is the covariance matrix. The factor analysis model represents $\boldsymbol{Z}$ as linearly dependent on a few latent factors ($\boldsymbol{F}$) given by the following equation

$$\boldsymbol{Z} - \boldsymbol{M} = k\boldsymbol{F} + \epsilon$$

The k given in the above equation represents factor loading matrix given by

$$k = \begin{bmatrix} k_{11} & \cdots & k_{1n} \\ \vdots & \ddots & \vdots \\ k_{q1} & \cdots & k_{qn} \end{bmatrix}$$

where k_{rs} represents r^{th} variable and s^{th}factor and n is the number of factors.

The factor scores thus obtained from the analysis are of the following form (Johnson and Wichern, 2014)

$$\widehat{f_s} = \begin{bmatrix} \frac{1}{\sqrt{\widehat{\lambda_1}}} \widehat{e_1}(Z_s - \bar{Z}) \\ \vdots \\ \frac{1}{\sqrt{\widehat{\lambda_n}}} \widehat{e_n}(Z_s - \bar{Z}) \end{bmatrix}$$

where $\widehat{f_s}$ is the factor score and $(\widehat{\lambda_s}, \widehat{e_s})$ are the estimated eigenvalue and eigenvector pairs.

Based on the factor scores obtained for each batsman or bowler, we arrive at their performance index.

6 Limitations and model testing

Uncontrollable variables in cricket like weather, coin-toss result, and field conditions can impact the performance of a player, but were not considered due to its inability to be quantified. Subsequently, ball-by-ball data needs to be extracted for IPL and T20I games for testing the model.

References

[1] Ahmed, F., Jindal, A., Deb, K. (2012), Multi-objective optimization and decision making approaches to cricket team selection. Appl. Soft Comput. doi: 10.1016/j.asoc.2012.07.031

[2] Amin, G. R., and Sharma, S. K. (2014). Measuring batting parameters in cricket: A two-stage regression-OWA method. *Measurement*, *53*, 56-61.

[4] Damodaran, U. (2006), Stochastic dominance and analysis of ODI batting performance: The Indian cricket team, 1989–2005. J. Sports Sci. Med. 5, 503–508

[5] Davis, J., Perera, H., and Swartz, T. B. (2015). A simulator for Twenty20 cricket. *Australian & New Zealand Journal of Statistics*, *57*(1), 55-71.

[6] Das, N. R., Ghosh, S., Mukherjee, I., and Paul, G. (2023). Adoption of a ranking based indexing method for the cricket teams. *Expert Systems with Applications*, *213*, 118796.

[7] Gupta, K. (2022). An integrated batting performance analytics model for women’s cricket using Principal Component Analysis and Gini scores. Decision Analytics Journal, 4, 100109.

[8] Johnson, R. A., and Wichern, D. W. (2002). Applied multivariate statistical analysis.

[9] Kamble, A., Rao, R., Kale, A., Samant, S. (2011), Selection of cricket players using analytical hierarchy process. Int. J. Sports Sci. Eng. 5, 207‑212

[10] Lemmer, H. H. (2002). The combined bowling rate as a measure of bowling performance in cricket. South African Journal for Research in Sport, Physical Education and Recreation, 24(2), 37-44.

[11] Lemmer, H. H. (2011a). The single match approach to strike rate adjustments in batting performance measures in cricket. Journal of sports science & medicine, 10(4), 630.

[12] Lemmer, H. H. (2011b). Performance measures for wicket keepers in cricket. South African Journal for Research in Sport, Physical Education and Recreation, 33(3), 89-102.

[13] Prakash, C. D., and Verma, S. (2022). A new in-form and role-based deep player performance index for player evaluation in T20 cricket. Decision Analytics Journal, 2, 100025.

[14] Premkumar, P., Chakrabarty, J. B., & Chowdhury, S. (2020). Key performance indicators for factor score based ranking in One Day International cricket. IIMB Management Review, 32(1), 85-95.

[15] Saikia, H., Bhattacharjee, D., and Lemmer, H. H. (2012). A double weighted tool to measure the fielding performance in cricket. International Journal of Sports Science & Coaching, 7(4), 699-713.

[16] Saikia, H., Lemmer, H., and Bhattacharjee, D. (2017). Quantify the fielding performance in cricket via bayesian approach. MOJ Sports Medicine, 1(4), 77-83.

[17] Saikia, H., Bhattacharjee, D., Mukherjee, D. (2019). Decision Making in Cricket: The Optimum Team Selection. Cricket Performance Management: Mathematical Formulation and Analytics, 201-231.

[18] Sharma, S. K. (2013). A factor analysis approach in performance analysis of T-20 cricket. Journal of Reliability and Statistical Studies, 69-76.

[19] Singh, G., Bhatia, N., Singh, S (2011), Fuzzy cognitive maps based cricket player performance evaluator. Int. J. Enterp. Comput. Bus. Syst. 1, 1‑15

[20] Swartz, TB, Gill, PS and Muthukumarana, S (2009), 'Modelling and simulation for one-day cricket,' Canadian Journal of Statistics, 37(2), pp. 143_160, doi:10.1002/cjs.10017

A machine learning-based approach to analyse player performance in T20 Cricket Internationals

Yash Jadwani*, James Denholm-Price**, Gordon Hunter***

School of Computer Science and Mathematics, Kingston University, KT1 2EE, UK
* yash.jadwani1998@gmail.com
** J.Denholm-Price@kingston.ac.uk
*** G.Hunter@kingston.ac.uk

Abstract

Cricket is one of the most-followed sports around the World. T20 is a short version of the game, growing in popularity over the past 20 years due to high profile tournaments such as the Indian Premier League. There is much demand for analysis of player performance, but traditional measures of this – batting and bowling averages, strike and economy rates - have limitations. We created a novel role-based performance metric using machine learning, allowing comparisons between players with similar roles in different teams. Using *ESPNCricinfo* data on T20 international matches, we calculated these new and the traditional performance metrics. Clustering was used to find “natural” classes of player types, a Random Forest classifier employed to identify the features most indicative of each cluster then PCA used to obtain the new performance indicator. Finally, we compared the results of our novel approach to the classical player performance metrics and player classifications provided by experts.

1 Introduction

T20 cricket is a fast-paced and exciting format of the game that demands players to score runs quickly and take wickets within a short span of time. However, the current ranking system for T20 cricket is not a fair and accurate assessment of T20 players. This is because the current system relies on traditional metrics, such as players’ career batting and bowling averages, strike rates and economy rates, without considering the distinct roles players undertake within a T20 team. Sports analytics has benefited greatly from the development and acceptance of machine learning methods (Deep et al., 2016). The traditional player performance metrics can help with these comparisons, but it has been well-documented (e.g. Attanayake and Hunter 2015) that each of these has limitations. This paper proposes novel performance metrics and a ranking system that takes into account player roles. The proposed system would assign different weights to different batting and bowling roles, based on the importance of those roles to a T20 team. This would allow for a more accurate assessment of T20 players, as it would take into account the different ways that players can contribute to a team. The proposed ranking system would be a valuable addition to the analytics framework for T20 cricket. It would foster a fairer evaluation process by capturing the diverse ways players can contribute to a team's success. This improvement would be beneficial to all stakeholders, including players, coaches, the media, and sports fans.

2 Related previous work

Traditional methods of comparing cricket performances rely on batting and bowling averages. However, these metrics have limitations as they oversimplify the evaluation process and don't account for the different roles played by players. For example, one of the few players to “average” over 100 in an English First Class season was the Australian bowler W.A. “Bill” Johnston in 1953, who scored 102 runs in 17 innings, but “averaged” 102.00 due to only being out once! In the context of shorter format cricket, where the primary objective for batters is to score runs quickly and for bowlers to concede as few runs as possible per over, traditional metrics such as batting and bowling averages may not provide a complete picture and “strike rates” and “economy rates” are frequently used instead. However, these also have their shortcomings, For example, the Indian fast bowler Zaheer Khan had the apparently excellent batting strike rates of 73.46 and 130.00 in one-day and T20 internationals respectively, but averages only 12.00 in the former and 6.50 in the latter. Although he scored runs quickly, he didn't score many. It is important to recognize that players have different roles within the team, such as opening batsmen are expected to score more runs than specialist bowlers, who are expected to bowl economically. To assess player performances in shorter format cricket, we should use multiple metrics to consider these various aspects of the game and each player's role in the team.

The cricket commentator and former Australian player Eddie Cowan (2011), following a suggestion by former Australia batting star Mike Hussey, suggested using "magic numbers" that find the sum of a batter's average and strike rate to evaluate their efficiency. Attanayake and Hunter (2015) instead proposed combining runs scored (or conceded), wickets lost (or taken) and balls faced (or bowled) in a multiplicative model, identifying previous metrics as special cases of this. Damodaran (2006) presented a Bayesian technique to deal with not-out scores in cricket as an alternative to batting average. Lemmer (2002) proposed a bowling method called CBR, combining traditional metrics using a formula: CBR = 3[1/Bowling Average + 1/Economy + 1/Strike Rate]. In subsequent papers, Lemmer analyzed bowling and batting performance, including T20 internationals, where new metrics were needed. He developed performance indicators and a formula to evaluate batters, considering their scores when out and not out. Spencer et al. (2016) used k-means clustering and Random Forest to cluster player/team profiles in the Australian Football League.

Deep et al. (2016) created DPI, a machine learning based approach to evaluate T20 players. It uses traditional metrics like economy, strike rate, and average, as well as new indicators like hard hitter, finisher, and running between the wickets (RWB). Deep et al. (2022) proposed DPPI, a more complex approach that considers player's form and role in the team. It uses predefined KPIs and adds new ones like Boundary index and Big Innings index.

3 Role-based player performance analysis in T20Is

The Player Performance Model developed by McHale et al. (2012) has proven successful in assigning a single score to players, irrespective of their specialty, based on their contributions to winning performances in the premier league and championship, the top two divisions of English football. Our study aimed to achieve the following objectives:

- Build a statistical index devoid of subjective opinions and understandable.
- Compare and evaluate players based on their respective roles within the team.
- Strike a balance between model simplicity and complexity.
- Prioritize runs for batters and wickets for bowlers when creating Key Performance Indicators (KPIs), as these represent the primary objectives for each player.

In our study, we conducted preliminary analysis, exploratory data analysis, and outlier removal to extract Key Performance Indicators (KPIs) for batters and bowlers. The assignment of player roles was based on the derived KPI values, without considering specific roles during the KPI establishment phase.

1	Boundaries Per Ball	Total boundaries (4s + 6s) / Total balls faced
2	Boundary Index	Total boundaries (4s + 6s) / Total innings batted
3	Finishing Index	Number of times batter remained not out / Total innings
4	Runs Without Boundary Index	Number of runs scored without boundaries / Total innings
5	Big Match Index	(2 * Centuries + half centuries) / Total innings batted

Table 1: Extracted KPIs for Batters

1	Balls Bowled per Innings	Number of balls bowled / Total innings bowled in
2	Wickets Index	Number of wickets taken / Total innings bowled in
3	Big Impact Index	Number of times bowler took 3 wickets or more / Total innings
4	Short Impact Index	Wickets taken without Big Impact Innings / Total innings
5	Runs Index	Number of Runs Conceded by bowler / Total innings bowled in

Table 2: Extracted KPIs for Bowlers

We cleaned the data and applied normalization techniques to ensure unbiased clustering. K-means clustering was used to group players with similar roles. This generated a target vector that was used for further steps. To determine the importance of each characteristic in relation to player roles, classification algorithms such as one-vs-all along with Random Forest were used. The feature importance values were multiplied by the normalized KPIs to derive Role-Based scores using Machine Learning (RBML) using a Random Forest classifier.

Additionally, Principal Component Analysis (PCA) was conducted on the normalized data, considering the correlations between the KPIs. The coefficients of the first component were used as feature importance values, which were multiplied by the normalized KPIs for comprehensive performance evaluation.

4 Data used and design of experiments

This study includes both active and retired players. We have used the dataset of 4000 entries from ESPNCricinfo, 2000 each for batters and bowlers. Data was gathered starting with the first T20 International match and continuing through April 17th, 2022. Preliminary metrics (PM) calculated for batters and bowlers using formulas by Basevi et al. (2007). PM for batters was (Average * Strike rate) / 100, and for bowlers, it was (Average * Economy) / 6. Prelim Rankings (PR) were assigned based on criteria such as PM values and runs for batters, and PM values and wickets for bowlers. The rankings included categories like Best, Good, Average, and Poor.

Criteria	Ranking
PM > 30 and runs >= 500	Best
PM > 30 and runs < 500	Good
PM between 20 and 30	Good
PM between 10 and 20	Average
PM < 10	Poor
Runs < 100	Poor

Table 3: Prelim Ranking Criteria for Batters

Criteria	Ranking
PM < 30 and Wickets >= 25	Best
PM < 30 and Wickets < 25	Good
PM between 30 and 40	Good
PM between 40 and 65	Average
PM > 65	Poor
Wickets < 5	Poor

Table 4: Prelim Ranking for Bowlers

4.1 Results and findings

After ranking the players, we extracted key performance indicators (KPIs) and applied feature scaling to mitigate possible machine learning algorithm bias. Subsequently, we utilized K-means clustering to determine player roles, treating these roles as the target vector for classification algorithms. To determine the importance of each KPI for each role, we employed a one-vs-all approach alongside a random forest classifier.

We identified four distinct batter roles: Specialist Batters (SB), Finishers (F), Floaters (FL), and Lower Contribution Batters (LCB). SB players excel in all KPIs except the finishing index, F players demonstrate quick scoring abilities while remaining not out, FL players exhibit high averages, good strike rates, and significant impact in big innings, while LCB players, with limited opportunities, focus on concluding innings strongly.

KPIs (Batters)	Feature importance is determined to identify the key features for each role			
	Specialist Batters (SB)	Finishers (F)	Floaters (FL)	Lower Contribution Batters (LCB)
Average	0.266	0.084	0.171	0.152
Strike Rate	0.065	0.212	0.075	0.383
Boundary Per Ball	0.061	0.166	0.074	0.199
Boundary Index	0.244	0.149	0.276	0.161
Finishing Index	0.027	0.158	0.096	0.027
Runs Without Boundary Index	0.085	0.176	0.186	0.054
Big Match Index	0.251	0.055	0.122	0.025

Table 5: Role-based feature importance for batters

Four distinct roles were identified for bowlers: Specialist Bowlers (SB) excel in all key performance indicators (KPIs), Short Performance Bowlers (SPB) have high averages and strike rates but concede more runs and have a lower wicket index than SB, Impact Bowlers (IB) maintain better economy rates than SPB while effectively restricting runs but bowl fewer overs than SB, and Lower Contribution Bowlers (LCB) bowl fewer overs than expected and take fewer wickets in comparison.

KPIs (Bowlers)	Feature importance is determined to identify the key features for each role			
	Specialist Bowlers (SB)	**Short Performance Bowlers (SPB)**	**Impact Bowlers (IB)**	**Lower Contribution Bowlers (LCB)**
Average	0.092	0.128	0.213	0.072
Strike Rate	0.092	0.06	0.104	0.061
Economy	0.038	0.122	0.082	0.052
Balls Bowled per Innings	0.204	0.233	0.141	0.399
Wicket Index	0.279	0.141	0.104	0.156
Big Impact Index	0.027	0.013	0.005	0.01
Short Impact index	0.199	0.085	0.073	0.145
Runs Index	0.068	0.218	0.278	0.105

Table 6: Role-based feature importance for bowlers

After that, we applied Principal Components Analysis (PCA) to verify the performance of clustering, reducing the data dimensionality, and obtained a performance indicator by multiplying the coefficient values of the first component with the KPIs. Although not role-based, this performance indicator allows us to evaluate the success of our supervised and unsupervised learning experiments.
The scatter plots below show the PCA components and group the points based on their assigned roles from clustering. For batters, higher values on the First component indicate better batters. For bowlers, lower values on the First component indicate better bowlers. Both plots effectively represent all the clusters, demonstrating the success of K-Means clustering in grouping the points.

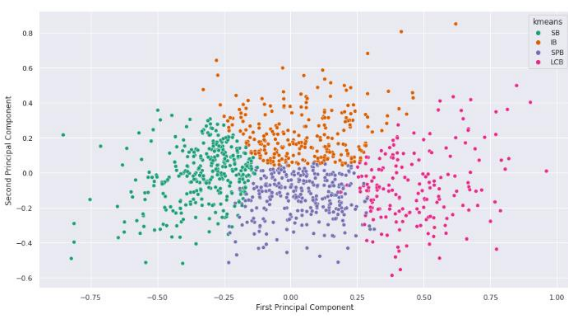


Figure 1: PCA for Batters

Figure 2: PCA for Bowlers

In summary, the scatter plot provides a visual representation of the role-based performance for both batters and bowlers, showcasing the effectiveness of K-Means clustering in grouping the points and highlighting different aspects of performance in cricket.

Player Name (Country)	Role given by K-means clustering	Preliminary Score	Prelim Rank	RBML Scores
M Hayden (AUS)	Specialist Batter	73.874	Good	2.903
Virat Kohli (IND)	Specialist Batter	70.900	Best	2.494
RA Jadeja (IND)	Finisher	27.036	Good	0.227
MK Pandey (IND)	Floater	55.897	Best	1.015
KC Sangakkara (SL)	Floater	37.539	Best	0.983
A Rashid (IND)	Specialist Bowler	27.479	Best	0.62
T Natarajan (IND)	Specialist Bowler	22.123	Good	0.771
LS Livingstone (ENG)	Short Performance Bowler	23.235	Good	0.621
Yuvraj Singh (IND)	Short Performance Bowler	20.968	Best	0.60
M Theekshana (SL)	Impact Bowler	29.377	Good	0.673
Sohail Tanvir (PAK)	Impact Bowler	32.214	Good	0.653

Table 7: Some players, along with their roles, RBML scores, and preliminary ranks

5 Discussion, conclusions and future work

Before clustering, the dataset was shuffled, and a subset of 98 batters and 82 bowlers was selected. This subset included Prelim Ranks (Best, Good, Average, Poor). Among the 'Poor' batters, 32 were classified as LCB, 12 as Finishers, 5 as Floaters, and 1 as SB. None of the 'Good/Best' batters were LCB, but there were more SB batters in the 'Good' ranks than 'Best'. Notably, ED Silva, initially ranked 'Poor', was classified as SB, suggesting the need for a broader performance evaluation. Among the 'Good' bowlers, 22 were SB, 8 were SPB, 7 were LCB, and 6 were IB. Out of the 'Best' bowlers, 9 were SB, while 1 each were SPB and IB. This highlights the importance of role-based classification over traditional rankings for bowlers, and that traditional metrics such as average, economy, and strike rate should not be the only performance indicators for a player.

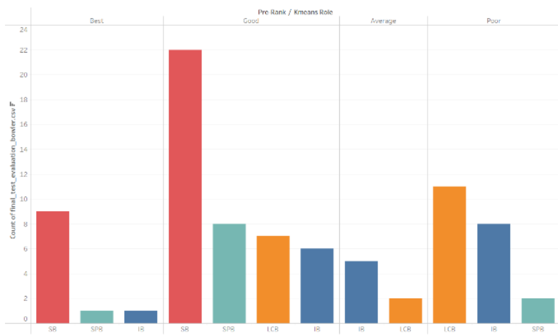


Figure 3: Results for Batters by category

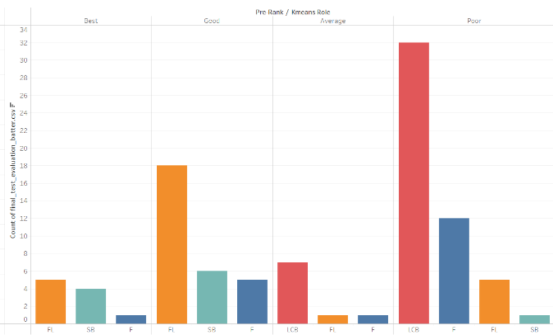


Figure 4: Results for Bowlers by category

In conclusion, the integration of k-Means clustering, classification algorithms, PCA, and role-based ranking provides valuable insights into player performance in T20 cricket. Furthermore, by comparing RBML/PCA ranks with roles assigned through clustering using PR rank and scores, we found that a player's performance cannot be solely determined by their average, strike rate, runs scored, or wickets taken. This approach offers a more comprehensive understanding of player capabilities and performance in the dynamic and fast-paced nature of T20 cricket.

Player evaluation in sports analytics is a growing field. This study has considered role-based clustering and predicted the feature importance of KPIs. However, recent form and pitch conditions can also play a role in player performance. In the future, the model developed by this study will remain valid, and one should be able to improve its performance by obtaining more data.

References

[1] Attanayake, D. and Hunter, G. , 2015. *Probabilistic Modelling of Twenty-Twenty (T20) Cricket : An Investigation into various Metrics of Player Performance and their Effects on the Resulting Match and Player Scores.* In Proceedings of 5th International Conference on Mathematics in Sport (Kay, Owen, Halkon and King., Eds.), Loughborough, U.K., June 2015.

[2] Cowan, E., 2011. *What's so great about a batting average* ? (On-line) https://www.espncricinfo.com/story/ed-cowan-on-more-meaningful-statistics-in-cricket-543061

[3] Spencer, B., Morgan, S., Zeleznikow, J. and Robertson, S. 2016. *Clustering team profiles in the Australian football league using performance indicators*, in: The 13th Australasian Conference on Mathematics and Computers in Sport, Melbourne, Australia, July 2016

[3] Basevi, T. and Binoy, G., 2007. *The world's best Twenty20 players*. [online] ESPNCricinfo. Available at: <https://www.ESPNCricinfo.com/story/the-world-s-best-twenty20-players-311962>

[4] Damodaran, U., 2006. *Stochastic dominance and analysis of ODI batting performance : The Indian cricket team 1989-2005*. Journal of Sports Science & Medicine, Vol. 5, pp. 503-508.

[5] Deep Prakash, C. and Verma, S., 2022. *A new in-form and role-based Deep Player Performance Index for player evaluation in T20 Cricket*. Decision Analytics Journal, 2, p.10002

[6] Deep Prakash, C., Patvardhan, C. and Singh, S., 2016. *A new Machine Learning based Deep Performance Index for Ranking IPL T20 Cricketers.* International Journal of Computer Applications, 137 (10) 5(2), pp. 42-49. Doi : 10.5120/ijca2016908903

[7] Lemmer, H., 2002. *The combined bowling rate as a measure of bowling performance in cricket*. South African Journal for Research in Sport, Physical Education and Recreation, 24(2).
[8] McHale, I., Scarf, P. and Folker, D., 2012. *On the Development of a Soccer Player Performance Rating System for the English Premier League*. Interfaces, 42 (4), pp. 339-351.

Analyzing the profitability of the football transfer market using network science and machine learning

Csaba Medgyes*

* Budapest University of Technology and Economics: csaba.medgyes.work@gmail.com

Abstract

The transfer fees in football have skyrocketed in the last 20 years, making the transfer market a valuable source of income for clubs, to the extent that several teams base their business plan on maximizing profits from player transfers. However, little research has been conducted on football's transfer market from a network science perspective. In this work, the market is considered as a time-varying network where the nodes are the football clubs (or leagues) and the links represent the transfer relationship between them with the transfer fees as weights. We collected more than 250 000 transfers from season 2010/11 until season 2021/22 considering the top-level leagues of 32 countries and constructed a large, time-varying football transfer network. After visualizing the network and investigating its most important structural characteristics, we use machine learning methods to predict the profitability of a transfer by combining network attributes and player attributes.

1 Introduction

The research on the transfer market from a network science perspective was initiated by (Lee et al. 2015). The analysis examined the top teams as nodes in the network and the links represented the transfers, with their weights corresponding to the transfer values. The main objective of the research was to investigate the distribution of weights and strengths within the network. The network was constructed with clubs as nodes and directed edges represented transfers in (Liu et al. 2016). The network role of a club was determined and the relationship between network properties and club functionality (profitability and average performance) was then assessed. The aim of (Bond et al. 2018) was to examine which countries were strong and influential in the transfer market. The network was constructed with countries as nodes, and the links represented the trading relationships between them. The study provided valuable insights into the market roles of emerging countries such as China, Russia, Brazil, and Turkey. The underlying hypothesis of (Füresz 2018) was to establish a relationship between a club's activity in the transfer market and the profit generated from those transfers. The clubs were represented as nodes in the network, and the transfers were links with associated weights, and centrality measures were used to explore the correlation patterns within the network. In their study (Li et al. 2019) constructed the network as the vertices were clubs, and the weight of each link represented the number of transfers between them. The analysis focused on studying the in-degree and out-degree distributions of the vertices, as well as the distribution of in- and out-strengths (sum of in- or out weights) among them. Furthermore, centrality analysis (betweenness and closeness centrality) was conducted to examine the centrality measures of the vertices in the network. The following study (Bond et al. 2020) focused on the topological characteristics of the European loan network, analyzing global and local network measures. The vertices of the network represented European teams, and the directed edges indicated the presence of loan transactions between them, with the weight of each edge corresponding to the number of loans. The primary objective was to determine whether these topological properties provided insights into any hidden structure within the European loan network among teams. The study (Xu 2021) explored several endogenous and exogenous measures where the network was constructed with the clubs as vertices, and the edges represented transfers between them, with the weights being the frequency of transfers. The first question addressed in the study was how the endogenous measures contributed to the formation of the network which aimed to understand the influence of factors such as reciprocity, popularity, and closure on the network's development. The second question focused on how the exogenous measures influenced the formation of the network which aimed to assess the impact of factors such as league type, match performance, and participation in the UEFA Champions League on the network's formation. The fundamental question examined in the study (Hoey et al. 2021) was the extent to which the transfer market redistributes revenues between large and small teams based on market size. To address this question, the researchers collected the official financial statements of each team and estimated market size using a regression model. The main concept of (Füresz & Rappai 2022) revolved around the idea that certain managers and agents have access to information about a particular transfer before it becomes known to investors, and how this affects the shares of the respective teams. The event study method was employed to examine whether there was any information leakage (abnormal stock movement) surrounding these transfers. Most recently (Wand 2022) the network structure was formed with clubs as the vertices and transfer transactions as the edges, with the weights representing the transfer values. The study examined various network properties, including degree distribution, disparity (heterogeneity), clustering coefficient, and small-world coefficient.

This paper aims to broaden the scope of analysis by including a larger number of leagues. We will examine 32 first-division leagues spanning from the 2010/2011 season to the 2021/2022 season. We will investigate whether the previous structural characteristics hold true when we expand our analysis to 32 leagues. Section 3 will employ machine learning techniques to detect profitability, taking into account the teams' network roles as potential influential factors. Section 4 will delve into analyzing the profit network of the transfer market.

2 Analyzing football's transfer network

In our project, we have expanded the number of leagues included. We gathered transfer data from 32 leagues, utilizing the information available on www.transfermarkt.com. All the scripts and datasets are available and public, which can be found in the following Github repository[1].

In this project, we analyzed the transfer market from the 2010/11 season until the 2021/22 season. We collected data from Transfermarkt and from the following nation's first-division leagues: Argentina, Austria, Belgium, Brazil, Bulgaria, China, Croatia, Czech Republic, Denmark, England, France, Germany, Greece, Hungary, Italy, Japan, Mexico, Netherlands, Norway, Poland, Portugal, Romania, Russia, Scotland, Serbia, Slovakia, Spain, Sweden, Switzerland, Turkey, Ukraine, United States. During our analysis we did not separate the summer and winter transfer windows, we joined them into one season. Since we did not have data about when the transfers actually happened we had to assume that the transfer happened at the beginning of the summer transfer window (10th June), so the market value before the transfer attribute has been calculated using this date. When considering transfer fees, we factored inflation in by dividing the fees using Eurostat data.

In exploratory data analysis, mainly trivial associations were observed. Due to COVID-19, many clubs experienced negative revenue and thus, there were seasons where teams did not spend as much money. Despite this, the number of transfers rose, however, the same trend was not observed for the average transfer fees of the seasons. The average transfer fee has dramatically risen in the last decade, nearly tripling in value, while the transfer values have not kept up with this trend. This demonstrates that the substantial investments made by clubs in the past decade have increased competition in the market for potential talent. Based on the positions and transfer fees or values, it is a well-known fact that players who play in attacking positions generally command higher average transfer fees and values, which has been confirmed.

[1] `https://github.com/csabamedgyes/Football_Transfer_Network`

When analyzing transfer fees and values across different leagues calculating basic averages can distort the results, so we incorporate the market importance factor of the league which is the ratio of the transfers in the given league compared to the total number of transfers. The English Premier League had a higher average buying price compared to the selling price, making it a buyer league, whereas France has a much lower buying price relative to the selling price. Therefore, in order to distinguish between buyer and seller leagues, we calculated the following:

$$\overline{X_{\mathrm{L}}^{\mathrm{b}}}\frac{|X_{\mathrm{L}}^{\mathrm{b}}|}{\sum_{i\in\mathrm{L}}|X_i^{\mathrm{b}}|}-\overline{X_{\mathrm{L}}^{\mathrm{s}}}\frac{|X_{\mathrm{L}}^{\mathrm{s}}|}{\sum_{i\in\mathrm{L}}|X_i^{\mathrm{s}}|} \tag{1}$$

where $X_{\mathrm{L}}^{\mathrm{b}}$ is the buying, $X_{\mathrm{L}}^{\mathrm{s}}$ is the selling transfer fee of that league. By using this the top 5 buyers league were England, China, Italy, Russia, and the United States, the seller ones were Portugal, Brazil, Netherlands, Argentina, and France.

We conducted an analysis of the Football Transfer Network, which is characterized by a set of nodes representing teams, and directed edges representing transfer relations between them, with the weight of each edge being the sum of the transfers between the teams. Our analysis focused on examining the network as a whole, as well as a time-varying network, wherein we examined the evolution, variations in parameters and alterations in characteristics.

The network structurally acted the expected way: in and out-degree distributions of the network appear to follow power-law distributions. The rich club phenomenon (rich club refers to a subset of nodes with high degrees that exhibit a tendency to establish connections amongst themselves) can be observed, with a high level of average clustering. From this, we can conclude that the expected results remain the same when extending our domain to 32 leagues. On a local level, the central teams are football's powerhouses: Manchester City, Chelsea, Barcelona, and Real Madrid which will change in Section 4 when the edge weight will be the profit. By treating it as a dynamic network and examining global measures, we can determine that structurally it operates in a similar manner, with no noteworthy alterations in its parameters.

3 Profit prediction using machine learning and network science

In this Section, our goal is to make predictions about profitability using Machine Learning. We will accomplish this by incorporating the network measures of both the buyer and seller teams. Our aim is to try to demonstrate that the attributes of the network measures have a noteworthy impact on predicting the profitability of a transfer.

We opt for binary classification due to the limited data available for monitoring player performances. The process of generating a target variable is relatively simple: suppose a player transferred from Club A to Club B, then from Club B to Club C, and so on. If the difference between the amount that Club B sold the player to Club C and the amount that Club B bought the player from Club A is positive, we assign a label of 1 to the transfer when the player moved from Club A to Club B, as it was a profitable transfer. If the difference is non-positive, we assign a label of 0, and if the buyer club has not sold the player yet, we assign a value of NaN.

Given our objective is to predict profitability using basic attributes and network measures of the buyer and seller teams, we have decided to calculate 6 centrality measures for each team and for each year. After we calculated all teams' network measures for that given year we inserted them into the next year's transfers. Upon further investigation, we observed that the 12 attributes (6 for the buyer and 6 for the seller) exhibit high correlations with each other. In order to mitigate this, we perform dimensionality reduction using Principal Component Analysis (PCA) separately for the buyer and seller measures. By reducing the 6 dimensions into one, we obtain the final attribute of the Buyer (or Seller) Network Measure.

Attribute	Attribute Type
Country (buyer)	category
Country (seller)	category
Age (at transfer)	float64
Position	category
Transfer value (before transfer)	float64
Mönchengladbach Height (cm)	float64
Nationality	category
Foot	category
Transfer Fee or type	float64
Buyer Network Measure	float64
Seller Network Measure	float64
Target	Binary

Table 1: Final Dataset's Attributes and types for Machine Learning

	Random Forest	AdaBoost	Light Gradient Boost
Precision	0.71898	0.7026	0.7297
Accuracy	0.90066	0.8965	0.9026
ROC-AUC	0.89291	0.8959	0.8889
F1	0.79113	0.7872	0.7916

Table 2: Model Evaluations

The final dataset (with all NaN excluded) consists of 15832 transfers and Table 1 shows its columns and corresponding data types. It is worth noting that the target attribute was imbalanced, as the distribution of the labels between 0 and 1 were 0.736:0.263 respectively.

In order to avoid leakage, which is the use of information in the model training process which would not be expected to be available at prediction time and to ensure better evaluation and explanation we did not split our dataset randomly into a standard 80%-20% train-test split. Instead, we divided our dataset into a training set and a test set based on the season of the transfers. The training set contained all transfers until the 17/18 season, and the test set contained all transfers from the 18/19 season. This resulted in a split of 77.236% for the training set and 22.764% for the test set.

After Hyperparameter Tuning (for precision, since it is selected as the primary performance measure because it is expensive for a club to invest in a player that is predicted to be profitable but turns out to be non-profitable) the results of the best-performing models can be found at the Table 2. We can conclude that the Light Gradient Boost outperforms the other two. The area under the precision-recall curve which is a suitable measure for imbalanced binary classification problems with rare positive instances was 0.7957, which was the highest among the three models. Therefore, we can conclude that the model can accurately identify positive instances while keeping a low false positive rate.

In LightGBM, split feature importance measures the frequency of a feature being used to split data. Features with higher split counts are considered more important. Gain feature importance calculates the overall reduction in loss function achieved by using a feature for data splitting in each tree node. It sums up the improvements in the loss function across all trees. The transfer fee remains the most crucial attribute in terms of split importance, but it is not a dominant factor in this model since the buyer and seller network measures are the other two significant attributes. In terms of gain importance, the transfer fee, transfer value, and age outperform the other two attributes.

	Team
1	Monaco
2	AS Roma
3	LOSC Lille
4	Chelsea
5	Barcelona
6	Sevilla FC
7	Atlético Madrid
8	Borussia Dortmund
9	Juventus
10	Benfica
11	FC Porto
12	Fiorentina
13	Olympique Lyon
14	Everton
15	Sampdoria
16	SSC Napoli
17	FC Internazionale
18	RB Salzburg
19	Valencia
20	Liverpool

(a) Weighted In Degree

	Team
1	Santos FC
2	Chelsea
3	LOSC Lille
4	Benfica
5	Rennes
6	FC Internazionale
7	Udinese Calcio
8	Sampdoria
9	FC Porto
10	Standard Liège
11	Internacional
12	Partizan
13	Real Madrid
14	SM Caen
15	KRC Genk
16	Olympique Marseille
17	Saint-Étienne
18	Paris Saint-Germain
19	RSC Anderlecht
20	Monaco

(b) Weighted Out Degree

	Teams
1	Juventus
2	Chelsea
3	VfL Wolfsburg
4	AS Roma
5	Valencia
6	Everton
7	Liverpool
8	SSC Napoli
9	Real Madrid
10	Monaco
11	Benfica
12	FC Internazionale
13	LOSC Lille
14	Atlético Madrid
15	OGC Nice
16	Udinese Calcio
17	Zenit St. Petersburg
18	Sassuolo
19	Manchester City
20	Bournemouth

(c) Eigenvector Centrality

Table 3: Weighted Profit In, Out Degree and Eigenvector Centrality (Top 20)

It is worth noting that transferring a player for free does not guarantee a label of 1, as the player may leave the team for free, resulting in a label of 0, so we need to restrict the test set to non-zero Transfer Fees. In this case, LightGBM is still the best classifier out of the three. The precision remained the same which is a positive outcome since we optimized for Precision during Hyperparameter Tuning.

4 Analyzing football's profit network

Using the previously defined target variable we can define football's profit network, where the teams are the nodes, and the edge weights are the realized profit (target). Although the network globally has the same characteristics as the transfer network, the central teams changed. In this Section, we will identify the central clubs within the profit network.

Based on how to network is created, we can infer that the weighted in-degree centrality measures the effectiveness of teams in transferring profitable players, while the weighted out-degree centrality identifies the teams from which players should be transferred in order to maximize our chances to realize profit. Monaco, AS Roma, and LOSC Lille were the top three teams that excelled in generating profits through their transfer strategies over the years (Table 3). Notably, some of the well-known football powerhouses, such as Chelsea and Barcelona, also appeared on the list, mainly due to their practice of spending substantial amounts on multiple transfers to acquire top talents. Based on the Weighted Out Degree, we can identify the top 20 teams from which players should be transferred in order to maximize our chances to realize a profit. This knowledge can be highly valuable for football clubs, since if a rumor surfaces that a particular team is interested in transferring a player, it may be a smart move to enter the competition for that player.

The Eigenvector Centrality reveals the extent to which a team's neighbors were successful in generating profits. If a team appears on this list but not on the other two, it could imply that their transfer strategy may not be the most effective for realizing profits. For instance, Everton is an example of a team that appears on the Eigenvector Centrality table but not on the other tables, suggesting that their transfer strategy may not be optimal for maximizing profits.

Considering closeness and betweenness centrality and their close association with weighted shortest paths, these measures enable us to identify the top-performing teams in terms of how quickly or efficiently profit can flow from the clubs in the network, and which teams act as a bridge or between other nodes in the network. In addition to the teams mentioned earlier, 1. FSV Mainz 05, Bayer 04 Leverkusen, and Borussia Mönchengladbach also appear, showing that German teams also intersect these paths of profit (sequence of profitable transfers).

Acknowledgements

The author would like to express their appreciation to Roland Molontay for helpful discussions and guidance, which significantly contributed to the completion of this project.

References

Lee, S., Hong, I., Jung, W.-S. (2015) *A Network approach to the transfer market of European football leagues*. New Physics: Sae Mulli, 65(4), 402-409.

Liu, X. F., Liu, Y.-L., Lu, X.-H., Wang, Q.-X., Wang, T.-X. (2016) *The anatomy of the global football player transfer network: Club functionalities versus network properties*. PloS One, 11(6), e0156504.

Bond, A. J., Widdop, P., Chadwick, S. (2018) *Football's emerging market trade network: Ego network approach to world systems theory*. Managing Sport and Leisure, 23(1-2), 70-91.

Fürész, D. I. (2018) *Correlation between profitability and transfer activity in European football*. Croatian Review of Economic, Business and Social Statistics, 4(2), 15-22.

Li, M.-X., Zhou, W.-X., Stanley, H. E. (2019) *Network analysis of the worldwide footballer transfer market*. EPL (Europhysics Letters), 125(1), 18005.

Bond, A. J., Widdop, P., Parnell, D. (2020) *Topological network properties of the European football loan system*. European Sport Management Quarterly, 20(5), 655-678.

Fürész, D. I., Rappai, G. (2022) *Information leakage in the football transfer market*. European Sport Management Quarterly, 22(3), 419-439.

Xu, Y. (2021) *The formation mechanism of the player transfer network among football clubs*. Soccer & Society, 22(7), 704-715.

Hoey, S., Peeters, T., Principe, F. (2021) *The transfer system in European football: A pro-competitive no-poaching agreement?*. International journal of industrial organization, 75, 102695.

Wand, T. (2022) *Analysis of the football transfer market network*. Journal of Statistical Physics, 187(3), 1-12.

Spatial statistics of images from players point of view in sport environments

Adam Y. Shavit* and Peter E. Leopold**

* Psychology department, Hunter College, CUNY

** BioAnalyte Inc. Portland, Maine.

Abstract

Perception of the visual space around the body is crucial for playing sports. First, to orient oneself relative to gravity, and then to interact with the environment. This work examines the visual spatial statistics available in images from the point of view of the observer in sports-related environments. We focus on the real component of the Fourier decomposition of the scene, on power analysis, and the orientation spectrum at different spatial frequencies. We discuss the literature on spatial cues for estimating the Visual Perceived Eye level, and the Visually Perceived Vertical. We vary the type of sport venue, in order to generate and later analyze the spatial statistics of these scenes. This work will also describe an algorithm for analyzing the effect of varying the orientation of the point of view in the roll and pitch rotations on the statistics of the image with implications for egocentric space perception.

1 Introduction

Athletes make split-second decisions based on what they see on a playing field or court. In an effort to model perceived (2D) visual fields for 3D perception using low- or no-parameter linear (fast) algorithms, we described an algorithm based on level sets of power spectra of 2D-FFTs. In this iteration of the development, we are able to perform basic image classification using 2D-FFT power spectra. Our image classification method can be compared to the modern convolutional neural network approach which uses 50 layers (for example, ResNet-50) of many-feature 2D-convolutional, many-feature 2D-pooling and dense neural layers requiring millions of parameters for basic image classification. Conversely, this paper uses a single relatively arbitrary parameter: the relative intensity of the power spectrum 2D level set.

This paper will analyze the power spectrum in images of sport venues and other image categories. The aim is to connect the literature on two research topics: The first focuses on the statistics of natural images and images affected by human activity, and on employing the techniques of power spectrum analysis. The second focuses on the effect of the visual field, and the orientation of inducers on spatial perception and judgments of space relative to the body.

The focus is on examining the orientation spectrum of images from the point of view of the observer. The analysis includes images of large soccer arenas, and basketball courts, in order to compare and contrast the visual information such as spatial frequency and orientation available to players with that available in natural visual environments, and architecture in urban environments. We will discuss these findings from the point of view of the effect of the visual field on perception and spatial judgements.

Egocentric space perception is the perception of the spatial relation between one's own point of view and the elements as well as the entire environment. The stability of egocentric space perception is the result of a multisensory integration (Matin & Li, 1995) of visual, vestibular, and proprioceptive cues (location of limbs in space).

Knill et al. (1998), Burge et al (2016), Hillis et al. 2004, and Chang et al. (2019) analyze how humans and ideal observers use local cues as well as cues pooled across the scene to estimate surface slant. Todd, 2005 describes the effect of field of view, judgements of slant depend on the frequencies and the size of the field of view. Field, 1987, as well as Simoncelli & Olshausen, 2001 describes how the visual system responds to natural and artificial visual environments.

2 Methods

We produced a 2D FFT power spectrum. The absolute magnitude of the real elements is shifted to locate 0 frequency in the center of the 2D FFTspectrum , logarithmically scaled and then linearly scaled relative to the max intensity of the 2D FFT power spectrum. The contours plot (Torralba & Oliva, 2003) represents the frequencies at different percentages of the energy levels (level sets), and the result is a topographical map of the orientations at different power or energy levels. There are 4 bands in the lower left image in the panels of Figure 2, and from these another smoothed contour is isolated at 99.5% level of energy.
Visual inspection of the power spectra of the different image categories allows us to identify the distribution of power spectra across different frequency bands unique to each category.

The analysis can continue by creating a variety of measures to quantify the differences between image categories in the power spectra. One metric that was helpful in our analysis is the ratio of curve length over a 180 degree arc to the area inside the curve, where the curve is specified in polar coordinates; This quantifies the roundness of the curve or the uniformity of frequencies across the spectrum of orientations. Another useful feature is the presence of bumps in the curve. These indicate an intensity peak in one or several orientations. Therefore, the second metric analyzed is the mean absolute change or mean absolute radio-angular acceleration $E|d^2r/d\theta^2|$ of the annulus.

3 Results

The results of the analysis outlined in the method are visualized in Figures 1 and 2. These include 2D-FFT images.

Natural images manifest a more round power spectrum annulus and are low in both measurements of ratio of curve length to area under the curve as well as the mean absolute radial-angular acceleration of the annulus. Figure 2 provides a scatterplot of these measurements against each other. Architectural photos have greater ratios, and sport photos have a variety of measurements.

These differences between images from different categories open the possibility of further classification and analysis. Figure 2 showed, in our limited sample of images, how (Nature with a capital N is the publication) images of natural settings cluster closely together, but away from urban and sport images despite a variety of similar characteristics. This is an interesting feature of diverse pictures including football, soccer field, kayak, basketball and more.

4 Discussion

Power spectrum analysis allows to differentiate between image categories according to metrics on the power spectrum.

The original Fourier spectrum as well as the histogram of orientations provided in Figure 1 are hard to interpret because they contain many features that are derived from the properties of the images themselves rather than the visual features of interest. For example, waves created from the image wrapping on itself in the Fourier analysis and artifacts introduced by the camera such as the focus of the image.

Therefore, the investigation proceeded with a power spectrum analysis and inspection of the equal energy curves. Natural environments are characterized by a more well rounded power spectrum. This is consistent with prior literature on the topic (Torralba, & Oliva, 2003; Van der schaaf, Van Hateren and associates 1998a, 1998b).

This current work adds to this literature by characterizing sport related images as artificial images and distinct from natural images. Additionally, describing the visual environment can lead to a better understanding of behavior in sport settings (Ripoll, 1995)

Future directions: With a larger sample of images and with additional information about the orientation of the camera, it may be possible to identify statistical relationships between measures of the power spectrum and image classification. This work may involve comparing the parameters of the power spectrum, analyzing the shape of the power spectrum distribution, or using specific metrics such as the slope of the power spectrum, ratio of properties, or metrics of shape as in our results. Figure 3 shows a start of this promising direction.

With a larger sample, our plan is to look for additional relationships in the image power spectrum. It will also be possible to run a multi-class logistic regression analysis to help classify image categories to sport-specific and other types of visual environments. Alternatively, using machine supervised learning methods, one can readily use a classifier, or look for k-neareast neighbors to create a classification map on the 2D space of the arc length-area ratio and expected absolute radioangular acceleration metrics.

In a more ambitious project, one can look for relationships across the time dimension by analyzing a continuous recording or a series of still images from the view of an observer. This can be done with a cyclopean vision style head camera attached to the athlete.

While this approach has many advantages over convolution neural networks for low resolution classification and intuition building, it may still make sense to train a neural network to predict the location and orientation of the camera from a test image after learning with a training set of images from different viewpoints in the same location. The CNN can be used to identify the orientation of the gravity vertical or the elevation of eye level. This approach would benefit from identifying measures in the power spectrum that correspond to roll-pitch-yaw of the image in 3D-space as well as to identifying the axis of bilateral symmetry and the vanishing point created by the retinal centric projective geometry of rolling or pitching the eyes (Matin, & Li, 1995; Shavit, A. Y.,2009).

Acknowledgements

The first author would like to thank Prof. Dana Sylvan. Thank you for your mentorship, for encouraging me to apply to MathSport, and for supporting me with an honorarium from the Hunter College Foundation.

References

[1] Burge, J., McCann, B.C. and Geisler, W.S., 2016. Estimating 3D tilt from local image cues in natural scenes. *Journal of Vision*, *16*(13), pp.2-2.

[2] Chang, T.Y., Thompson, L., Doudlah, R., Kim, B., Sunkara, A. and Rosenberg, A., 2020. Optimized but not maximized cue integration for 3D visual perception. *ENeuro*, *7*(1).

[3] Field, D.J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), pp.2379-2394.

[4] Hillis, J.M., Watt, S.J., Landy, M.S. and Banks, M.S., 2004. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12), pp.1-1.

[5] Knill, D.C., 1998. Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research, 38*(11), pp.1655-1682.

[6] Matin, L. and Li, W., 1995. Multimodal basis for egocentric spatial localization and orientation. *Journal of Vestibular Research: equilibrium & orientation*, *5*(6), pp.499-518.

[7] Olshausen, B.A. and Field, D.J., 1996. Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images. *Nature*, *381*, pp.607-609.

[8] Ripoll, H., Kerlirzin, Y., Stein, J.F. and Reine, B., 1995. Analysis of information processing, decision making, and visual strategies in complex problem solving sport situations. *Human Movement Science*, *14*(3), pp.325-349.

[9] Peven, J.C., Chen, Y., Guo, L., Zhan, L., Boots, E.A., Dion, C., Libon, D.J., Heilman, K.M. and Lamar, M., 2019. The oblique effect: The relationship between profiles of visuospatial preference, cognition, and brain connectomics in older adults. *Neuropsychologia*, *135*, p.107236.

[10] Shavit, A.Y., 2009. *Employing a straight array of parallel identical lines to study the effect of visual orientation on the visual perception of egocentric space*. Columbia University.

[11] Simoncelli, E.P. and Olshausen, B.A., 2001. Statistical properties of natural images. *Annual Review of Neuroscience*, *24*.

[12] Todd, J.T., Thaler, L. and Dijkstra, T.M., 2005. The effects of field of view on the perception of 3D slant from texture. *Vision Research*, *45*(12), pp.1501-1517.

[13] Torralba, A. and Oliva, A., 2003. Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), p.391.

[14] Van der Schaaf, V.A. and van Hateren, J.V., 1996. Modeling the power spectra of natural images: statistics and information. *Vision Research*, *36*(17), pp.2759-2770.

[15] Van Hateren, J.H. and van der Schaaf, A., 1998. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), pp.359-366.

[16] van Hateren, J.H. and Ruderman, D.L., 1998. Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1412), pp.2315-2320.

[17] Yang, Z. and Purves, D., 2003. Image/source statistics of surfaces in natural scenes. *Network: Computation in Neural Systems*, *14*(3), p.371.

[18] Yang, Z. and Purves, D., 2003. A statistical explanation of visual space. *Nature Neuroscience*, *6*(6), pp.632-640.

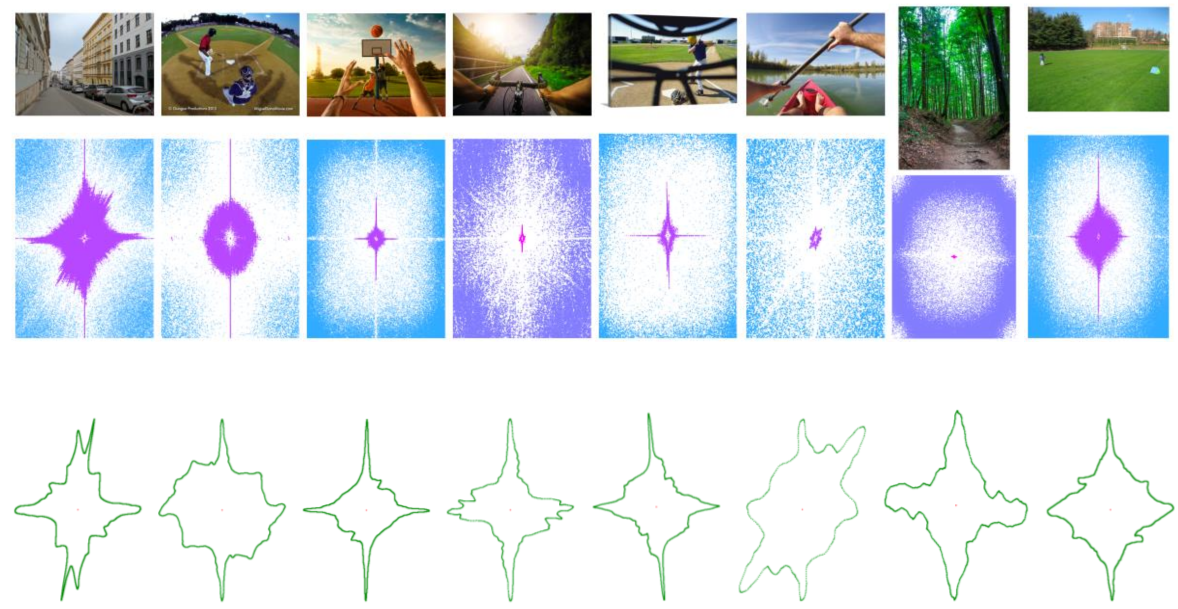


Figure 1. The analysis follows the power analysis in Torralba & Oliva (2003). Eight Panels are presented depicting the power analysis of a single image. The top row contains the image, The middle row the original Fourier Transform $F(u, v)$ in 4 energy bands. The third row contains the smoothed annulus of 99.5% high energy contour.

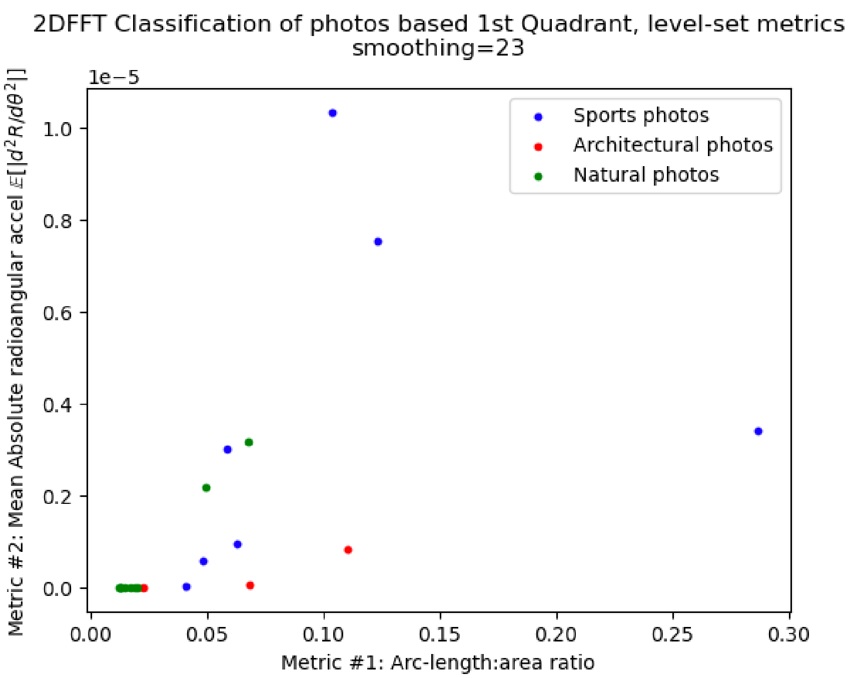


Figure 2. A scatterplot of power spectrum measurements. Mean Absolute Radio-angular Acceleration is plotted against the ratio of Arc-Length to Area ratio for the annulus of individual power spectrums for the images analyzed in this study. Images from different categories scatter in a nonrandom fashion. Natural images are low in both measurements. Architectural photos have greater ratios.

ISBN 978-963-503-941-8