BDSports, a network of people interested in Sports Analytics

http://bodai.unibs.it/bdsports/

BDSports is a project developed by the Big&Open Data Innovation Laboratory (BODal-Lab) of the University of Brescia, Italy.

Scientific coordinators of the project: Paola Zuccolotto and Marica Manisera

OUR FRIENDS: Research Centers and Laboratories
- Data Methods and Systems Statistical Laboratory, University of Brescia (Italy)
- KPA Group - Insights through Analytics, Raanana (Israel)
- Strategic Research in Sport Center, Moscow (Russia)
- AUEB Sports Analytics Group, Athens University of Economics and Business (Greece)

OUR FRIENDS: Products & Services Platforms for Sports Analytics
- StatBasket, Como (Italy)
- Mθagonom, Brescia (Italy)
- Quant4Sport, Torino (Italy)
- Football Intelligence, Siena (Italy)
- Math4Sport, Milano (Italy)

Description of the project:
In the last decades, quantitative thinking in sports has gained a rapidly growing interest. This is reflected by the scientific production on this theme and also by the publication of collections of statistical analyses applied to data from a wide range of sports, including football, basketball, volleyball, baseball, ice hockey and many others. This research project is designed to set up a unique collaboration of experts interested in sport analytics both from a scientific and a practical point of view. The goal is to create a network able to facilitate contacts and joint research initiatives. Specifically, the project will organise events, promote special issues in scientific journals, share ideas and data in order to publish scientific and non-scientific papers, collaborate with teams in various sports by supplying them analytics and apply for research grants.

The data scientists’ expertise covers a wide range of quantitative tools in the fields of statistical modelling, multivariate data analysis, data mining, algorithmic modelling and machine learning.

Main topics:
- Basic statistics and more complex analytics of a match or a competition
- Performance analysis (of teams, players, individual athletes)
- Identification of success factors and optimal game strategies
- Forecasting
- Sport Psychology (group dynamics, interpersonal relations, social-cognitive processes, leadership, mental toughness, personality, coping strategies, ...)
- Market analysis for sport marketing
- Financial assessment of sports clubs and sports related projects
Agenda:

• Basketball analytics: state of the art
• Basketball datasets
• Case studies:
  • CS1: new positions in basketball
  • CS2: scoring probability when shooting under high-pressure conditions
  • CS3: performance variability and teamwork assessment
  • CS4: sensor data analysis
• Concluding remarks
Basketball Analytics

- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

Basketball Analytics

Official Statistics

Sport Analytics Services

Scientific Research

BDsports

Big Data analytics in sports

Marica Manisera
Paola Zuccolotto

University of Brescia, Italy
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis
Our analyses often integrate machine learning tools and experts’ suggestions.

- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis
As concerns basketball, several statistical techniques have been applied to analyze data with a great variety of different aims, ranging from simply depicting the main features of a game by means of descriptive statistics (Kubatko et al., 2007) to the investigation of more complex problems, such as forecasting the outcomes of a game or a tournament (West et al., 2008; Loeffelholz et al., 2009; Brown et al., 2010; Gupta, 2015; Lopez and Matthews, 2015; Ruiz and Perez-Cruz, 2015; Yuan et al., 2015; Manner, 2016), analysing players’ performance (Page et al., 2007; Cooper et al., 2009; Piette et al., 2010; Fearnhead and Taylor, 2011; Ozmen, 2012; Page et al., 2013; Deshpande and Jensen, 2016), studying the network of players’ pathways from the in-bounds pass to the basket (Skinner, 2010) and their spatial positioning (Shortridge et al., 2014), or identifying optimal game strategies (Annis et al., 2006).
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball
• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball

• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis

Stats

CS1

www.espn.com/nba
stats.nba.com
www.fiba.com
Leagues

...
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball

Data

Big Data

Sensor

Data

CS4

Stats

CS1

play-by-play

CS2 – CS3

• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
**Role revolution: towards a new meaning of positions in basketball**

Federico Bianchi\textsuperscript{a}, Tullio Facchinetti \textsuperscript{a}, and Paola Zuccolotto\textsuperscript{b}

\textsuperscript{a}University of Pavia, via Ferrata, 1, 27100 Pavia, Italy
\textsuperscript{b}University of Brescia, c.d.q S. Chiara 50, 25122 Brescia, Italy

**Motivation:** The existing positions - often defined a long time ago - tend to reflect traditional points of view about the game and sometimes they are no longer well-suited to the new concepts arisen with the evolution of the way of playing.

**Aim:** describing new roles of players during the game, by means of the analysis of players' performance statistics with data mining and machine learning tools.
• Basketball Analytics: state of the art
• Basketball datasets
• **CS1: new positions in basketball**

• «Key-players» training set
→ 7-dimensional SOM

• clusterization of the SOM output layer into a proper number of groups by means of a **fuzzy clustering algorithm**
• Basketball Analytics: state of the art
• Basketball datasets
• **CS1**: new positions in basketball
• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
- Basketball Analytics: state of the art
- Basketball datasets
- **CS1:** new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis
**Motivation**: Basketball players have often to face high-pressure game conditions. To be aware of the overall and personal reactions to these situations is of primary importance to coaches.

**Aim**: To develop a model describing the impact of some high-pressure game situations on the probability of scoring and to assess players' personal reactions.
HiGH-Pressure Game Situations:

• when the shot clock is going to expire (SHOT.CLOCK)
• when the score difference with respect to the opponent is small (Sc.DIFF)
• when the team, for some reason, has globally performed bad during the match, up to the considered moment (MISS.T)
• when the player missed the previous shot (MISS.PL)
• the time to the end of quarter (TIME)
• type of action (POSS.TYPE, 24” or 14” extratime)
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

### Dataset

<table>
<thead>
<tr>
<th>Competition Period Gender</th>
<th>A2ITA</th>
<th>Rio16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of matches</td>
<td>480</td>
<td>38</td>
</tr>
<tr>
<td>Number of teams</td>
<td>32</td>
<td>12</td>
</tr>
<tr>
<td>Number of players</td>
<td>438</td>
<td>144</td>
</tr>
<tr>
<td>Number of 2-point shots</td>
<td>33682 (48.3%, 50.9% Made)</td>
<td>3101 (47.9%, 52.2% Made)</td>
</tr>
<tr>
<td>Number of 3-point shots</td>
<td>21163 (30.4%, 34.1% Made)</td>
<td>1780 (27.5%, 33.8% Made)</td>
</tr>
<tr>
<td>Number of free throws</td>
<td>14843 (21.3%, 73.5% Made)</td>
<td>1589 (24.6%, 74.8% Made)</td>
</tr>
</tbody>
</table>

69688 6470
Data Mining Tools:

- **univariate** non-parametric regressions via kernel smoothing on the dependent variable Made (assuming values 1 and 0 according to whether, respectively, the attempted shot scored a basket or not)

- 1000 bootstrap samples of size $n_{boot} = 5000$ and $n_{boot} = 1000$ for the dataset A2ITA and Rio16, respectively.

  few univariate relationships detected - Just Shot.Clok and Miss.Pl
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball
• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis
Data Mining Tools:

- **CART** (Classification And Regression Trees), algorithm able to deal with multivariate complex relationships, also detecting interactions among predictors
- we transform numerical into categorical covariates in order to improve interpretability → combination of the results of a machine learning procedure and experts' suggestions
- pruning

Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis
focus on:

(1) the last 1-2 seconds of possession, very close to the shot clock buzzer sounding,
(2) games where the score difference is low, for example, between -4 and 4,
(3) the last 1-2 minutes of each quarter (especially the final quarter)
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball

• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis

<table>
<thead>
<tr>
<th>SHOT.CLOCK</th>
<th>early: SHOT.CLOCK &gt; 17</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>early-middle: 10 &lt; SHOT.CLOCK ≤ 17</td>
</tr>
<tr>
<td></td>
<td>middle-end: 2 &lt; SHOT.CLOCK ≤ 10</td>
</tr>
<tr>
<td></td>
<td>time-end: SHOT.CLOCK ≤ 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TIME</th>
<th>normal: TIME ≤ 500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>quarter-end: TIME &gt; 500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Miss.T</th>
<th>Bad: Miss.T ≤ 0.44 (25th percentile)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Medium: 0.44 &lt; Miss.T ≤ 0.56</td>
</tr>
<tr>
<td></td>
<td>Good: Miss.T &gt; 0.56 (75th percentile)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sc.DIFF</th>
<th>less than -15: Sc.DIFF ≤ -15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>between -15 and -5: -15 &lt; Sc.DIFF ≤ -5</td>
</tr>
<tr>
<td></td>
<td>between -5 and 1: -5 &lt; Sc.DIFF ≤ 1</td>
</tr>
<tr>
<td></td>
<td>between 1 and 6: 1 &lt; Sc.DIFF ≤ 6</td>
</tr>
<tr>
<td></td>
<td>more than 6: Sc.DIFF &gt; 6</td>
</tr>
</tbody>
</table>

focus on:

1. the last 1-2 seconds of possession, very close to the shot clock buzzer sounding,
2. games where the score difference is low, for example, between -4 and 4,
3. the last 1-2 minutes of each quarter (especially the final quarter)
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball

• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis

Very similar results with Rio 2016 data
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- **CS2**: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

**2-point shots**

- Shot type: 2P
- Shot Clock: time-end, middle-end
  - Shot Clock: time-end
  - Poss. type: 24sec
    - Sc. Diff: less than -15
      - 0.4678
      - 0.5501
  - Poss. type: 24sec
    - 0.6035
    - 0.6580
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball

• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball

free throws

Miss Pl: Made
0.7481 0.7115

• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
New Shooting Performance Measure:

Takes into account that shots attempted in different moments have different scoring probabilities

Performance of Player $i$

$$P_i(T) = \alpha \sum_{j \in J_T} (x_{ij} - \pi_{ij})$$

for shot type T (2P, 3P, FT)

$j$-th shot made (1)

or missed (0)

scoring probability of $j$-th shot according to CART
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

![Basketball Analytics Diagram]
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball
• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis
Further Research:

According to psychological studies, some athletes view the competitive situations as challenging, and others perceive the same situations as stressful and anxiety-provoking. For this reason, it may be difficult to statistically detect stressful situations from large datasets including several players, as the overall average performance may remain unchanged as a response to some players improving their performance and some others getting worse.

- Analysis of single players’ reactions to stressful game situations (propensity to shot and variation in scoring probability)
- Integration with psychological studies

University of Brescia, Italy
Motivation: Psychological studies have pointed out that typical performance is but one attribute of performance, but other aspects should be taken into account, in particular performance variability.

Aim: Assessment of players' shooting performance variability and investigation of its relationships with the team composition.
Performance Variability:

- Definition of a performance index based on the % of attempted shots that scored a basket and on the shooting intensity

\[
\tilde{\phi}_{ij} = \frac{1}{t_{ij}} \quad \phi_{ij} = \frac{\tilde{\phi}_{ij}}{\phi(m_{ij})}
\]

\[
E_{ij} = x_{ij} - p_{ij} \quad \psi_{ij} = \tilde{\phi}_{ij}E_{ij}
\]
Performance Variability:

• Fit **Markov Switching models** to the shooting performance index, in order to detect the (significant) presence of periods of good and bad performance.
Teamwork Assessment:

- determine influence of each teammate on the regime of good and bad performance
- display the significant relationships by means of graphical network analysis tools
- predict the best substitution at a given time
**CS4: sensor data analysis**

**Space-Time Analysis of Movements in Basketball using Sensor Data**

Marica Manisera *1, Rodolfo Metulini †1, and Paola Zuccolotto ‡1

1University of Brescia - Department of Economics and Management, Contrada Santa Chiara, 50, 25122, Brescia, Italy

**Aim:** A first approach to sensor data analysis in basketball (visualization tools, cluster analysis, future challenges)

In collaboration with MYagonism
Basketball Analytics: state of the art
Basketball datasets
CS1: new positions in basketball
CS2: scoring probability under high-pressure
CS3: performance variability and teamwork
CS4: sensor data analysis

**CS4: sensor data analysis**

Modelling the dynamic pattern of surface area in basketball and its effects on team performance

**Aim:** to study
- structural changes in the surface area
- associations between regimes and game variables
- the relation between the regime probabilities and the scored points
Visualization Tools

Spatio-Temporal Movements in Team Sports: A Visualization approach using Motion Charts.

Rodolfo Metulini (1)

(1) Department of Economics and Management, University of Brescia, Contrada Santa Chiara, 50, 25122 Brescia BS, Italy.
rodolfo.metulini@unibs.it.

A tool to display data recorded by tracking systems producing spatio-temporal traces of player trajectories with high definition and frequency

https://www.youtube.com/watch?v=aejyrDnqYVy
- Basketball Analytics: state of the art
- Basketball datasets
- CS1: new positions in basketball
- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

**Visualization Tools**

**James P. Curley**
Curley Social Neurobiology Lab website
(Psychology Department and Center for Integrative Animal Behavior, Columbia University, New York City)
Basketball Analytics: state of the art
Basketball datasets
CS1: new positions in basketball
CS2: scoring probability under high-pressure
CS3: performance variability and teamwork
CS4: sensor data analysis

Convex Hulls Analysis

<table>
<thead>
<tr>
<th></th>
<th>Average distance</th>
<th>Convex hull area</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>attack</td>
<td>defence</td>
</tr>
<tr>
<td>Min</td>
<td>5.418</td>
<td>2.709</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>7.689</td>
<td>3.942</td>
</tr>
<tr>
<td>Median</td>
<td>8.745</td>
<td>4.696</td>
</tr>
<tr>
<td>Mean</td>
<td>8.426</td>
<td>5.548</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>9.455</td>
<td>5.611</td>
</tr>
<tr>
<td>Max</td>
<td>10.260</td>
<td>11.640</td>
</tr>
</tbody>
</table>
Basketball Analytics: state of the art
Basketball datasets
CS1: new positions in basketball

CS2: scoring probability under high-pressure
CS3: performance variability and teamwork
CS4: sensor data analysis

Cluster Analysis
Basketball Analytics: state of the art
Basketball datasets
CS1: new positions in basketball

- CS2: scoring probability under high-pressure
- CS3: performance variability and teamwork
- CS4: sensor data analysis

Cluster Analysis + Multidimensional Scaling

BDsports
Big Data Analytics in Sports

Marica Manisera
Paola Zuccolotto
- University of Brescia, Italy
Basketball Analytics: state of the art
Basketball datasets
CS1: new positions in basketball

CS2: scoring probability under high-pressure
CS3: performance variability and teamwork
CS4: sensor data analysis

Cluster Analysis + MultiDimensional Scaling

<table>
<thead>
<tr>
<th>NA</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>10.71</td>
<td>23.53</td>
<td>47.83</td>
<td>0</td>
<td>20.83</td>
<td>31.25</td>
<td>20.23</td>
</tr>
<tr>
<td>2</td>
<td>0.77</td>
<td>0</td>
<td>9.15</td>
<td>0</td>
<td>1.85</td>
<td>2.08</td>
<td>8.33</td>
<td>2.89</td>
</tr>
<tr>
<td>3</td>
<td>31.54</td>
<td>42.86</td>
<td>0</td>
<td>8.7</td>
<td>44.44</td>
<td>20.83</td>
<td>18.75</td>
<td>20.23</td>
</tr>
<tr>
<td>4</td>
<td>6.15</td>
<td>3.57</td>
<td>1.96</td>
<td>0</td>
<td>7.41</td>
<td>0</td>
<td>10.42</td>
<td>1.16</td>
</tr>
<tr>
<td>5</td>
<td>0.77</td>
<td>3.57</td>
<td>16.99</td>
<td>17.39</td>
<td>0</td>
<td>0</td>
<td>16.67</td>
<td>8.09</td>
</tr>
<tr>
<td>6</td>
<td>27.69</td>
<td>7.14</td>
<td>18.95</td>
<td>0</td>
<td>1.85</td>
<td>0</td>
<td>0</td>
<td>43.93</td>
</tr>
<tr>
<td>7</td>
<td>15.38</td>
<td>21.43</td>
<td>3.92</td>
<td>4.35</td>
<td>18.52</td>
<td>0</td>
<td>0</td>
<td>3.47</td>
</tr>
<tr>
<td>8</td>
<td>17.69</td>
<td>10.71</td>
<td>25.49</td>
<td>21.74</td>
<td>25.93</td>
<td>56.25</td>
<td>14.58</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.12</td>
<td>85.88</td>
</tr>
<tr>
<td>14.09</td>
<td>85.91</td>
</tr>
<tr>
<td>26.07</td>
<td>73.93</td>
</tr>
<tr>
<td>15.38</td>
<td>84.62</td>
</tr>
<tr>
<td>57.98</td>
<td>42.02</td>
</tr>
<tr>
<td>85.07</td>
<td>14.93</td>
</tr>
<tr>
<td>39.53</td>
<td>60.47</td>
</tr>
<tr>
<td>39.08</td>
<td>60.92</td>
</tr>
</tbody>
</table>
Basketball Analytics: state of the art
Basketball datasets
CS1: new positions in basketball

Future Challenges:

- Integration with play-by-play data
- Integration with video and match analysis
- Integration with body metrics (body physiology tracking via “smart clothing” and/or body measurements)
- Integration with qualitative assessments

- Network analysis tools
- Spatio-temporal statistical models

- Addition of the other team’s data
- Addition of the ball’s position
• Basketball Analytics: state of the art
• Basketball datasets
• CS1: new positions in basketball
• CS2: scoring probability under high-pressure
• CS3: performance variability and teamwork
• CS4: sensor data analysis

Concluding ...

**True:**

• If people keep thinking that Statistics is merely PPG, AST, REB, ...
• If people don’t learn how Stats have to be interpreted (“Do not put your faith in what statistics say until you have carefully considered what they do not say.” W. W. Watt)

**False:**

• If modern approaches to basketball analytics are used
• If we are able to integrate analytics and technical experience
• If we are able to spread the culture of Statistics
REFERENCES

Download a (regularly updated) list of references at http://bodai.unibs.it/bdsports/basketball.htm

THANK YOU