Football Analytics

Hadi Sotudeh

January 2020
Agenda

• Introduction
• Datasets
• Applications
• Case Study 1
• Case Study 2
• Q & A
Introduction
Datasets
Event Data:

• On-ball actions
• ~2000 events per match
• Easy to query:
  • All the final third passes by Karimi
• Without context
Event Data Example:

```xml
<event>
  <id>11</id>
  <production_on>valde-jobq-a02.nexus.opta.net</production_on>
  <production_time>20170119T183125.932</production_time>
  <production_module>Opta Feed: XML: Soccer: F24</production_module>
  <game_id>953008</game_id>
  <away_team_name>Caen</away_team_name>
  <competition_id>24</competition_id>
  <competition_name>French Ligue 1</competition_name>
  <game_date>2017-01-18T19:00:00</game_date>
  <home_team_id>1459473765</home_team_id>
  <period_id>16</period_id>
  <time>0:0</time>
  <event_id>11</event_id>
  <type_id>34</type_id>
  <min>0</min>
  <sec>0</sec>
  <outcome>1</outcome>
  <timestamp>2017-01-18T18:25:98</timestamp>
  <qualifier_id>003556525</qualifier_id>
  <value>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30</value>
  <event_id>1542689574</event_id>
  <qualifier_id>30</qualifier_id>
  <value>9997, 203368, 42761, 11591, 12845, 44287, 194095, 86281, 60319, 45802, 50996, 7103, 97443, 6595, 199666, 58619, 33946, 115961, 40281, 192995, 136839, 87520895, 585487239</value>
  <event_id>1458083580</event_id>
  <qualifier_id>44</qualifier_id>
  <value>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30</value>
  <event_id>1542689574</event_id>
  <qualifier_id>227</qualifier_id>
  <value>0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0</value>
  <event_id>1697029712</event_id>
  <qualifier_id>130</qualifier_id>
  <value>11</value>
  <event_id>230429696</event_id>
  <qualifier_id>194</qualifier_id>
  <value>45802</value>
</event>
```
Assignment 1

France Ligue 1, 2016 – 2017, first half of the season:

tinyurl.com/ty6op7q
Tracking Data:

• Optical tracking cameras:
  • 90 minutes
    • 90*60 = 5400 seconds
  • 22 players + ball
  • 23 objects
  • 10 frames per second
    • =~ (5400*23*10 = 1,242,000) rows
  • 25 frames per second
    • =~ (5400*23*25= 3,105,000) rows
Tracking Data:

• GPS:

https://images.app.goo.gl/hkxWjWz8gUuixqEm9
SportVU Cameras

• Missile tracking

• 2010-2011 in NBA

• 2016-2017 in Ligue de Football Professional (France)
SportVU Cameras

- Two types:
  - Data
  - Video and Data

- 30,000 USD per camera

- 3 cameras per stadium
How do they work?

1. Tag the objects in the beginning to track:
   - Kernelized Correlation Filters (KCF)

2. Detect the jersey numbers and the ball in each frame:
   - YOLO: You Only Look Once
You Only Look Once: 
Unified, Real-Time Object Detection

Joseph Redmon*, Santosh Divvala†, Ross Girshick¶, Ali Farhadi*†
University of Washington*, Allen Institute for AI†, Facebook AI Research¶


Abstract

We present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to $448 \times 448$, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model’s confidence.
## Tracking Data Example

<table>
<thead>
<tr>
<th>PLAYER_ID</th>
<th>TEAM_FIXTURE</th>
<th>HALF</th>
<th>PLAYER_X_POSITION</th>
<th>PLAYER_Y_POSITION</th>
<th>TIME_VALUE</th>
<th>X_POS</th>
<th>Y_POS</th>
</tr>
</thead>
<tbody>
<tr>
<td>723245</td>
<td>Lyon</td>
<td>First Half</td>
<td>52.64</td>
<td>33.82</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>347689</td>
<td>Paris SG</td>
<td>First Half</td>
<td>42.41</td>
<td>33.71</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>346567</td>
<td>Paris SG</td>
<td>First Half</td>
<td>41.90</td>
<td>57.61</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>347763</td>
<td>Paris SG</td>
<td>First Half</td>
<td>49.40</td>
<td>52.81</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>348502</td>
<td>Paris SG</td>
<td>First Half</td>
<td>38.18</td>
<td>38.07</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>347858</td>
<td>Lyon</td>
<td>First Half</td>
<td>71.99</td>
<td>14.62</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>464719</td>
<td>Lyon</td>
<td>First Half</td>
<td>52.91</td>
<td>10.39</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>375843</td>
<td>Paris SG</td>
<td>First Half</td>
<td>33.51</td>
<td>36.63</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>437146</td>
<td>Paris SG</td>
<td>First Half</td>
<td>52.35</td>
<td>43.78</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
<tr>
<td>440664</td>
<td>Lyon</td>
<td>First Half</td>
<td>74.17</td>
<td>42.43</td>
<td>0.0</td>
<td>52.5</td>
<td>34.0</td>
</tr>
</tbody>
</table>

**Lyon v PSG**
Assignment 2

France Ligue 1, 2016 – 2017, 17 matches:

tinyurl.com/yx4vuznb
Applications
Scouting

“There are rich teams
There are poor teams
Then there's 50 feet of crap
And then there's us
It's an unfair game.”
Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos  
KU Leuven  
Leuven, Belgium  
tom.decroos@cs.kuleuven.be

Jan Van Haaren  
SciSports  
Amersfoort, Netherlands  
j.vanhaaren@scisports.com

Lotte Bransen  
SciSports  
Amersfoort, Netherlands  
l.bransen@scisports.com

Jesse Davis  
KU Leuven  
Leuven, Belgium  
jesse.davis@cs.kuleuven.be

ABSTRACT
Assessing the impact of the individual actions performed by soccer players during games is a crucial aspect of the player recruitment process. Unfortunately, most traditional metrics fall short in addressing this task as they either focus on rare actions like shots and goals alone or fail to account for the context in which the actions are performed. Our work introduces a new approach to valuing player actions that takes into account both the rarity of the action and the context in which it is performed. We demonstrate the effectiveness of our approach through experiments on real-world soccer data.

1 INTRODUCTION
How will a soccer player’s actions impact his or her team’s performances in games? This question is relevant for a variety of tasks within a soccer club such as player acquisition, player evaluation, and scouting. It is also important for the media and building fan engagement, as fans like nothing better than comparing players and their contributions to the team. To address this problem, we introduce a new approach to valuing player actions that takes into account both the rarity of the action and the context in which it is performed.
A game is a sequence of on-the-ball actions $[a_1, a_2, \ldots, a_m]$ where $m$ is the total number of actions.

- **StartTime**: the action’s start time,
- **EndTime**: the action’s end time,
- **StartLoc**: the $(x, y)$ location where the action started,
- **EndLoc**: the $(x, y)$ location where the action ended,
- **Player**: the player who performed the action,
- **Team**: the player’s team,
- **ActionType**: the type of the action (e.g., *pass*, *shot*, *dribble*),
- **BodyPart**: the player’s body part used for the action,
- **Result**: the result of the action (e.g., *success* or *fail*).
Given: game state $S_i = [a_1, \ldots, a_i]$;

**Estimate:** the probability of scoring and conceding in the near future for the home team $h$ and the visiting team $v$, which

$$P_{scores}(S_i, h) = P(goal(h) \in F^k_i | S_i)$$

$$P_{concedes}(S_i, h) = P(goal(v) \in F^k_i | S_i)$$

$$P_{scores}(S_i, v) = P(goal(v) \in F^k_i | S_i)$$

$$P_{concedes}(S_i, v) = P(goal(h) \in F^k_i | S_i)$$

where $F^k_i = [a_{i+1}, \ldots, a_{i+k}]$ is the sequence of $k$ actions that follow action $a_i$, and $k$ is a user-defined parameter.
Binary classification:

- Any machine learning algorithm that predicts a probability:
  - Logistic Regression, Random Forest, or Neural Network

- The probability estimates should be well-calibrated

\[
g([a_{i-2}, a_{i-1}, a_i]) = y_i
\]

\[ \Delta P_{\text{scores}}(a_i, x) = P_{\text{scores}}(S_i, x) - P_{\text{scores}}(S_{i-1}, x). \quad (1) \]

\[ \Delta P_{\text{concedes}}(a_i, x) = P_{\text{concedes}}(S_i, x) - P_{\text{concedes}}(S_{i-1}, x). \quad (2) \]
Experiments

• Two catboost models (scoring and conceding probabilities)

• Train the first model on:
  • the 2012/2013 through 2015/2016 seasons
  • produce the outcomes for the 2016/2017 season
  • ROC AUC: 0.76

• Train the second model on:
  • the 2012/2013 through 2016/2017 seasons
  • produce the outcomes for the 2017/2018 season
  • ROC AUC: 0.73

Barcelona’s goal against Real Madrid on December 23, 2017

VAEP (Valuing Actions by Estimating Probabilities)

\[ V(a_i, x) = \Delta P_{scores}(a_i, x) + (-\Delta P_{concedes}(a_i, x)) \]  

(3)

\[ rating(p) = \frac{90}{m} \sum_{a_i \in A_p^T} V(a_i), \]  

(4)

Results

At least 900 mins in the 2017/2018 English Premier League Season

(d) Top-10 players in terms of our VAEP player ratings

<table>
<thead>
<tr>
<th>( R_{vaep} )</th>
<th>Player</th>
<th>Rating</th>
<th>( R_g )</th>
<th>( R_a )</th>
<th>( R_{g+a} )</th>
<th>Market Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P. Coutinho</td>
<td>0.899</td>
<td>10</td>
<td>2</td>
<td>4</td>
<td>€ 140m</td>
</tr>
<tr>
<td>2</td>
<td>M. Salah</td>
<td>0.817</td>
<td>1</td>
<td>23</td>
<td>2</td>
<td>€ 150m</td>
</tr>
<tr>
<td>3</td>
<td>K. De Bruyne</td>
<td>0.641</td>
<td>72</td>
<td>4</td>
<td>15</td>
<td>€ 150m</td>
</tr>
<tr>
<td>4</td>
<td>E. Hazard</td>
<td>0.636</td>
<td>21</td>
<td>122</td>
<td>34</td>
<td>€ 150m</td>
</tr>
<tr>
<td>5</td>
<td>R. Mahrez</td>
<td>0.635</td>
<td>34</td>
<td>11</td>
<td>16</td>
<td>€ 60m</td>
</tr>
<tr>
<td>6</td>
<td>A. Martial</td>
<td>0.607</td>
<td>13</td>
<td>13</td>
<td>9</td>
<td>€ 60m</td>
</tr>
<tr>
<td>7</td>
<td>R. Sterling</td>
<td>0.579</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>€ 120m</td>
</tr>
<tr>
<td>8</td>
<td>P. Pogba</td>
<td>0.549</td>
<td>55</td>
<td>9</td>
<td>28</td>
<td>€ 80m</td>
</tr>
<tr>
<td>9</td>
<td>H. Kane</td>
<td>0.545</td>
<td>4</td>
<td>140</td>
<td>6</td>
<td>€ 150m</td>
</tr>
<tr>
<td>10</td>
<td>S. Heung-Min</td>
<td>0.539</td>
<td>19</td>
<td>36</td>
<td>17</td>
<td>€ 50m</td>
</tr>
</tbody>
</table>

Results

Born after 1/1/1997

2017/2018 season

(b) Young talents in the French, Dutch, and Belgian leagues.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Team</th>
<th>Age</th>
<th>Rating</th>
<th>Market Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D. Neres</td>
<td>Ajax</td>
<td>21</td>
<td>0.620</td>
<td>€ 25m</td>
</tr>
<tr>
<td>2</td>
<td>M. Mount</td>
<td>Vitesse</td>
<td>19</td>
<td>0.616</td>
<td>€ 4m</td>
</tr>
<tr>
<td>3</td>
<td>Malcom</td>
<td>Bordeaux</td>
<td>21</td>
<td>0.567</td>
<td>€ 40m</td>
</tr>
<tr>
<td>4</td>
<td>K. Mbappé</td>
<td>PSG</td>
<td>19</td>
<td>0.507</td>
<td>€ 200m</td>
</tr>
<tr>
<td>5</td>
<td>F. de Jong</td>
<td>Ajax</td>
<td>20</td>
<td>0.495</td>
<td>€ 60m</td>
</tr>
</tbody>
</table>

Results

(b) Playing style of replacement players for Ronaldo

Players in the 2017/2018 season who played at least 900 mins

(a) All players

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Contract</th>
<th>SciSkill</th>
<th>Potential</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kevin De Bruyne</td>
<td>28</td>
<td>2023-06</td>
<td>138.7</td>
<td>0.1</td>
<td>Attacking midfield</td>
</tr>
<tr>
<td>Roberto Firmino</td>
<td>28</td>
<td>2023-06</td>
<td>134.4</td>
<td>2.4</td>
<td>Centre forward</td>
</tr>
<tr>
<td>David Silva</td>
<td>33</td>
<td>2020-06</td>
<td>133.3</td>
<td>0.4</td>
<td>Attacking midfield</td>
</tr>
<tr>
<td>Georginio Wijnaldum</td>
<td>29</td>
<td>2021-06</td>
<td>131.8</td>
<td>1.4</td>
<td>Centre midfield</td>
</tr>
<tr>
<td>Dele Alli</td>
<td>23</td>
<td>2024-06</td>
<td>122.5</td>
<td>0.2</td>
<td>Attacking midfield</td>
</tr>
<tr>
<td>Jordan Henderson</td>
<td>29</td>
<td>2023-06</td>
<td>122.0</td>
<td>1.4</td>
<td>Centre midfield</td>
</tr>
<tr>
<td>Christian Eriksen</td>
<td>27</td>
<td>2020-06</td>
<td>118.8</td>
<td>3.2</td>
<td>Attacking midfield</td>
</tr>
<tr>
<td>N'Golo Kanté</td>
<td>28</td>
<td>2023-06</td>
<td>116.3</td>
<td>0.4</td>
<td>Centre midfield</td>
</tr>
</tbody>
</table>
Data scouting at PEC Zwolle

Validating existing targets

Suggest new targets
Assignment 3
Find **3 suitable candidates** for the **Right Back (RB)** position at your team (counter-attack based team)

**Conditions:**

- Under 25 (Young)
- Up to 750,000 Euro (Cheap enough!)
- Asian or African (Open new markets)
- Brilliant Future (Transfer and make a profit)

[ tinyurl.com/u7ybrqg ]
### Players.csv

<table>
<thead>
<tr>
<th>birthArea</th>
<th>birthDate</th>
<th>currentNationalTeamId</th>
<th>currentTeamId</th>
<th>firstName</th>
<th>foot</th>
<th>height</th>
<th>lastName</th>
<th>passportArea</th>
<th>role</th>
<th>shortName</th>
<th>weight</th>
<th>wyld</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netherlands</td>
<td>7/25/1998</td>
<td>671</td>
<td>23568</td>
<td>Thijmen</td>
<td>right</td>
<td>196</td>
<td>Nijhuis</td>
<td>Netherlands</td>
<td>Goalkeeper</td>
<td>T. Nijhuis</td>
<td>83</td>
<td>365439</td>
</tr>
<tr>
<td>Netherlands</td>
<td>8/12/1999</td>
<td>664</td>
<td>9</td>
<td>Matthijs</td>
<td>right</td>
<td>188</td>
<td>de Ligt</td>
<td>Netherlands</td>
<td>Defender</td>
<td>M. de Ligt</td>
<td>89</td>
<td>365443</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2/28/1998</td>
<td>670</td>
<td>10</td>
<td>Teun</td>
<td>left</td>
<td>183</td>
<td>Koopmeiners</td>
<td>Netherlands</td>
<td>Midfielder</td>
<td>T. Koopmeiners</td>
<td>71</td>
<td>365444</td>
</tr>
<tr>
<td>Netherlands</td>
<td>3/13/1998</td>
<td>670</td>
<td>32</td>
<td>Jay-Roy</td>
<td>right</td>
<td>183</td>
<td>Grot</td>
<td>Suriname</td>
<td>Forward</td>
<td>J. Grot</td>
<td>98</td>
<td>365445</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1/19/1999</td>
<td>670</td>
<td>11</td>
<td>Donyell</td>
<td>right</td>
<td>179</td>
<td>Malen</td>
<td>Suriname</td>
<td>Forward</td>
<td>D. Malen</td>
<td>78</td>
<td>365446</td>
</tr>
<tr>
<td>Austria</td>
<td>2/4/1998</td>
<td>9116</td>
<td>680</td>
<td>Maximilian</td>
<td>left</td>
<td>189</td>
<td>WAJFber</td>
<td>Austria</td>
<td>Defender</td>
<td>M. WAJFber</td>
<td>82</td>
<td>367503</td>
</tr>
</tbody>
</table>

### KPIs.csv

<table>
<thead>
<tr>
<th>wyld</th>
<th>accelerations</th>
<th>aerialDuels</th>
<th>aerialDuelsWon</th>
<th>assists</th>
<th>backPasses</th>
<th>corners</th>
<th>crosses</th>
<th>directRedCard</th>
<th>dribbles</th>
<th>forwardPasses</th>
<th>fouls</th>
<th>goals</th>
<th>headShots</th>
<th>interceptions</th>
<th>longPasses</th>
<th>losses</th>
<th>matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>149157</td>
<td>1</td>
<td>19</td>
<td>5</td>
<td>0</td>
<td>102</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>19</td>
<td>175</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>53</td>
<td>34</td>
<td>97</td>
<td>11</td>
</tr>
<tr>
<td>218554</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>216558</td>
<td>15</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>85</td>
<td>5</td>
<td>28</td>
<td>0</td>
<td>70</td>
<td>90</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>25</td>
<td>19</td>
<td>103</td>
<td>18</td>
</tr>
<tr>
<td>221036</td>
<td>1</td>
<td>17</td>
<td>10</td>
<td>0</td>
<td>30</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>80</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>35</td>
<td>59</td>
<td>9</td>
</tr>
<tr>
<td>221037</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>26</td>
<td>59</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>12</td>
<td>43</td>
<td>6</td>
</tr>
<tr>
<td>Player</td>
<td>Age</td>
<td>Nat.</td>
<td>Club</td>
<td>Market value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>--------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hans Hateboer</td>
<td>25</td>
<td></td>
<td></td>
<td>€15.00m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Denzel Dumfries</td>
<td>23</td>
<td></td>
<td></td>
<td>€13.50m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joël Veitman</td>
<td>27</td>
<td></td>
<td></td>
<td>€10.00m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timothy Fosu-Mensah</td>
<td>21</td>
<td></td>
<td></td>
<td>€7.00m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenny Tete</td>
<td>24</td>
<td></td>
<td></td>
<td>€7.00m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rick Karsdorp</td>
<td>24</td>
<td></td>
<td></td>
<td>€7.00m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
External Datasets

M. Taremi (ID: 241788) | FIFA 20 DEC 18, 2019
Mehdi Taremi  🇮🇷  CF LW  Age 26 (Jul 18, 1992)  6'2"  176lbs

- **Overall Rating**: 74
- **Potential**: 74
- **Value**: €6M
- **Wage**: €10K

### Real Overall Rating
- LS: 75
- ST: 75
- RS: 75
- LW: 73
- LF: 75
- CF: 75
- RF: 75
- RW: 73
- LAM: 73
- CAM: 73
- RAM: 73
- LM: 71
- LCM: 66
- CM: 66
- RCM: 66
- RM: 71
- LWB: 53
- LDM: 51
- CDM: 51
- RDM: 51
- RWB: 53
- LB: 50
- LCB: 46
- CB: 46
- RCB: 46
- RB: 50

### Attacking
- Crossing: 56
- Finishing: 74
- Heading Accuracy: 69
- Short Passing: 68
- Volleys: 71
- Dribbling: 77
- Curve: 71
- FK Accuracy: 43
- Long Passing: 47
- Ball Control: 78

### Movement
- Acceleration: 68
- Sprint Speed: 67
- Agility: 67
- Reactions: 74
- Balance: 60
- Strength: 72
- Long Shots: 70

### Skill
- Shot Power: 78
- Jumping: 70
- Stamina: 66
- Strength: 72
- Long Shots: 70

### Mentality
- Aggression: 31
- Interceptions: 17
- Positioning: 76
- Vision: 68
- Penalties: 61
- Composure: 74

### Defending
- Defensive Awareness: 33
- Standing Tackle: 37
- Sliding Tackle: 19

### Goalkeeping
- GK Diving: 9
- GK Handling: 11
- GK Kicking: 9
- GK Positioning: 15
- GK Reflexes: 8

---

https://sofifa.com
Talent detection in soccer using a one-class support vector machine

Jauhiainen Susanne1*, Forsman Hannele2, Äyrämö Sami1, and Kauppi Jukka-Pekka1

1 Faculty of Information Technology, University of Jyvaskyla, Finland
2 Eerikkilä Sports Institute, Tammela, Finland

Abstract — A large variety of different features are needed to succeed in soccer and therefore talent identification is a very complex process. In this study, unsupervised anomaly detection was used to get new insight into talent identification. The aim was to build an automatic ”talent detector” from a longitudinal player data set to detect those Finnish players that have signed a contract with an soccer academy abroad. One-class support vector machine (one-class SVM) was able to detect these players with perfect sensitivity. A portion of other players were also detected as ”talented” (AUC = 0.77). This may indicate that they also have potential to succeed abroad. On the other hand, the number of used variables was small due to high number of missing values, and it is likely that the specificity of the model can be improved once more data is obtained.
Pre-Match Analysis
Pre-Match Analysis

• Search in videos:
  • Find all moments that:
    • Central defenders were 30+ meters apart

https://twitter.com/KubaMichalczyk/status/1211291230140796928
Coding Data:

https://images.app.goo.gl/MovFXrnSDbcvyUKj9
Half-Time Break!
Challenges

1. Having a high data quality:

Challenges

2. Merging event data with tracking data

3. Adding context to event data:
   • Is a pass under pressure?
   • What are the pass options?

4. Writing queries for tracking datasets

5. Finding public work
Challenges

6. Creating new metrics
Case Study 1
Question

How does a football team possess the ball?

Event Data:

`period`: 2,
`timestamp`: "00:07:22.793",
`minute`: 52,
`second`: 22,
`type`: {
  "id": 16,
  "name": "Shot"
},
`possession`: 86,
`possession_team`: {
  "id": 780,
  "name": "Portugal"
},
`play_pattern`: {
  "id": 5,
  "name": "Other"
},
`off_camera`: false,
`team`: {
  "id": 780,
  "name": "Portugal"
},
`player`: {
  "id": 5207,
  "name": "Cristiano Ronaldo dos Santos Aveiro"
},
`location`: [108.0, 40.0],
`duration`: 0.68,
`related_events`: ["7a835a88-2a92-466d-9461-dea4d0f4f61f"],
`shot`: {
  "statsbomb_xg": 0.76,
  "end_location": [119.0, 42.2, 0.9],
  "outcome": {
    "id": 100,
    "name": "Saved"
  },
  "body_part": {
    "id": 40,
    "name": "Right Foot"
  },
  "type": {
    "id": 88,
    "name": "Penalty"
  },
  "technique": {
    "id": 93,
    "name": "Normal"
  }
}

https://scroll.in/field/884104/fifa-world-cup-iran-hold-portugal-to-1-1-draw-as-cristiano-ronaldo-misses-penalty

https://github.com/statsbomb/open-data
Ball possession process
<table>
<thead>
<tr>
<th>caseID</th>
<th>action</th>
<th>type</th>
<th>play_pattern</th>
<th>recipient</th>
<th>startTime</th>
<th>period</th>
<th>duration</th>
<th>endTime</th>
</tr>
</thead>
<tbody>
<tr>
<td>7557160</td>
<td>Pass</td>
<td>Free Kick</td>
<td>From Free Kick</td>
<td>Omid Ebrahimi</td>
<td>00:48:41.660</td>
<td>2</td>
<td>1.6</td>
<td>00:48:43.260000</td>
</tr>
<tr>
<td>7557160</td>
<td>Pass</td>
<td>None</td>
<td>From Free Kick</td>
<td>Sardar Azmoun</td>
<td>00:48:46.460</td>
<td>2</td>
<td>2.76</td>
<td>00:48:49.220000</td>
</tr>
<tr>
<td>7557160</td>
<td>Pass</td>
<td>None</td>
<td>From Free Kick</td>
<td>Saman Ghoddos</td>
<td>00:48:49.220</td>
<td>2</td>
<td>1.12</td>
<td>00:48:50.340000</td>
</tr>
<tr>
<td>7557160</td>
<td>Shot</td>
<td>Open Play</td>
<td>From Free Kick</td>
<td></td>
<td>00:48:51.753</td>
<td>2</td>
<td>0.107</td>
<td>00:48:51.860000</td>
</tr>
<tr>
<td>7557160</td>
<td>Block</td>
<td>None</td>
<td>From Free Kick</td>
<td></td>
<td>00:48:51.860</td>
<td>2</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>7557160</td>
<td>Goal Keeper</td>
<td>Shot Faced</td>
<td>From Free Kick</td>
<td></td>
<td>00:48:51.995</td>
<td>2</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>7557160</td>
<td>Ball Recovery</td>
<td>None</td>
<td>From Free Kick</td>
<td></td>
<td>00:48:53.380</td>
<td>2</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>7557160</td>
<td>Shot</td>
<td>Open Play</td>
<td>From Free Kick</td>
<td></td>
<td>00:48:53.993</td>
<td>2</td>
<td>0.44</td>
<td>00:48:54.433000</td>
</tr>
<tr>
<td>7557160</td>
<td>Goal Keeper</td>
<td>Shot Faced</td>
<td>From Free Kick</td>
<td></td>
<td>00:48:54.433</td>
<td>2</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>possession_team</th>
<th>team_action</th>
<th>player</th>
<th>body_part</th>
<th>start_X</th>
<th>start_Y</th>
<th>end_X</th>
<th>end_Y</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iran</td>
<td>Iran</td>
<td>Ramin Rezaeian</td>
<td>Right Foot</td>
<td>28.0</td>
<td>78.0</td>
<td>28.0</td>
<td>54.0</td>
<td>None</td>
</tr>
<tr>
<td>Iran</td>
<td>Iran</td>
<td>Omid Ebrahimi</td>
<td>Right Foot</td>
<td>35.0</td>
<td>51.0</td>
<td>86.0</td>
<td>53.0</td>
<td>None</td>
</tr>
<tr>
<td>Iran</td>
<td>Iran</td>
<td>Sardar Azmoun</td>
<td>Right Foot</td>
<td>84.0</td>
<td>40.0</td>
<td>97.0</td>
<td>39.0</td>
<td>None</td>
</tr>
<tr>
<td>Iran</td>
<td>Iran</td>
<td>Saman Ghoddos</td>
<td>Right Foot</td>
<td>97.0</td>
<td>45.0</td>
<td>101.0</td>
<td>45.0</td>
<td>Blocked</td>
</tr>
<tr>
<td>Iran</td>
<td>Portugal</td>
<td>Josabove Miguel da Rocha Fonte</td>
<td></td>
<td>20.0</td>
<td>36.0</td>
<td>20.0</td>
<td>36.0</td>
<td>None</td>
</tr>
<tr>
<td>Iran</td>
<td>Portugal</td>
<td>Rui Pedro dos Santos PatrÃ­cio</td>
<td></td>
<td>2.0</td>
<td>39.0</td>
<td>2.0</td>
<td>40.0</td>
<td>None</td>
</tr>
<tr>
<td>Iran</td>
<td>Iran</td>
<td>Mehdi Taremi</td>
<td>Left Foot</td>
<td>112.0</td>
<td>33.0</td>
<td>120.0</td>
<td>35.5</td>
<td>Off T</td>
</tr>
<tr>
<td>Iran</td>
<td>Iran</td>
<td>Mehdi Taremi</td>
<td>Left Foot</td>
<td>113.0</td>
<td>33.0</td>
<td>120.0</td>
<td>35.5</td>
<td>Off T</td>
</tr>
<tr>
<td>Iran</td>
<td>Portugal</td>
<td>Rui Pedro dos Santos PatrÃ­cio</td>
<td></td>
<td>4.0</td>
<td>45.0</td>
<td>5.0</td>
<td>45.0</td>
<td>None</td>
</tr>
</tbody>
</table>
Belgium

7 matches
Disco

- Activities: 100%
- Paths: 0%
Statistics

- 594 cases
- 295 variants
Set Pieces (sub-processes)

• Action + Type
Undesired cases
Undesired cases
## Undesired cases

### Cases (5)
- 8650119: 1 events
- 7536147: 1 events
- 7536161: 1 events
- 7584137: 1 events
- 7552129: 1 events

### Case with 1 events

<table>
<thead>
<tr>
<th>period</th>
<th>minute</th>
<th>second</th>
<th>duration</th>
<th>endTime</th>
<th>possession_team</th>
<th>team_action</th>
<th>player</th>
<th>player_position</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>73</td>
<td>55</td>
<td>00:28:57.320000</td>
<td>Belgium</td>
<td>Belgium</td>
<td>Jan Vertonghen</td>
<td>Left Back</td>
<td></td>
</tr>
</tbody>
</table>

Event: 01.01.
Undesired cases
Next Steps

• Redefine the process
Process Mining Meets Football! How Does a Football Team Possess The Ball On The Pitch?  Rudi 23 Oct

Finding the right perspective of the process is one of the challenges you can face when applying process mining. In most cases we already have an idea what we would expect of the process, but in some cases it's not so easy to find the right perspective to be able to get valuable insights.
Case Study 2
SPORTS ANALYTICS CHALLENGE
15 years ago

Offensive Organization

* Team organized in a pure 4x3x3. From last year have improved in the aggressiveness of their offensive transition but on the other hand are much weaker in that characteristic possession game and are much more dependable on the creativity of Messi and Ronaldinho. Although it's only 1st leg they will try to score at St. Bridge. A lot of mistakes in 1st phase build up but in 3rd phase and 4th phase deadly to finish all situations. Will constantly simulate free-kicks and penalties.
* Short or long build-up will depend on the amount of space and incentivation we give to them. Field is made big by the full-backs who open out as wide as possible. This positioning is dangerous in 1st phase as it leaves the central defenders with no support and makes the line of pass to the full-backs easy to intercept. Will use their typical combination to exit: 1) Deco short. 2) Gio wide. 3) Ball from Deco to Gio who frees Ronaldinho inside or 4) Ball back to Edmilson to organize.
* Long build-up is oriented to Eto'o and Ronaldinho. No real power to flick ball in depth and great possibility of us to win 1st ball and give continuity by win 2nd ball.

**Oleguer and Edmilson can be the ideal targets for high pressure**

Both with poor notions of time and space. With Edmilson is important to let him receive the ball first and then has he turns surprise him with pressure. With Oleguer his important to reduce the space as the ball is travelling to force him to a mistake. Transition!
* Always be ready for Gio's arrival from behind. Wants depth inside when Ronaldinho stays opened getting into a position to shoot or on the overlap if he's inside.
* With Xavi out the moments of pure possession and game domination are much shorter. Their possession now is under much more threat has there more players prone to make mistakes. In midfield Van Bommel plays simple organization passes but has power and intensity to cover a great radius of action supporting behind the line of the ball. Deco wants to penetrate with the ball (good target to pressure) but his most dangerous movement is the vertical switch when Eto'o short.
* When Ronaldinho provokes between the lines it's important to communicate with defensive midfielder because the positioning can be far away for our full back to control (if full-back follows all the way winger has to cover space outside because Gio will penetrate from behind). This momentum can be stopped by fouls.
* Messi very different than Giuly. Last year with Giuly more depth and width in attack. Messi is the contrary. First he has total freedom and even ends up on the opposite side to create 2vs1 with Ronaldinho. He wants to receive the ball early, linking phases of play by pure ball driving (mainly coming inside to his left foot). Brings creativity and risk to the game in 4th phase. Normally his 1vs1 is a simple touch on the ball to the side on the limit of the defenders intervention. Fouls!
* Attention on the switches of position: One Eto'o movement circular to the right side with Messi (Larsson vs Belis) coming inside on an aggressive diagonal-danger.
* See pag. 4 for Ronaldinho's pattern 1st time pass to Eto'o when ball is switching from right side to the left. Eto'o anticipates and reads the off-side line-depth.

In their 1st phase we can force mistakes or losses of possession by pressing Oleguer and Edmilson in the correct timing.
Pressing

**OLEGUER** and **EDMILSON**

can be the ideal targets for high pressure
• Challenge:
  • To identify pressing patterns in the last 3rd of the field

• Goal:
  • To develop automatic, scalable, and objective tool
  • To provide feedback to the coaching staff

• Approach:
  • To employ *Machine Learning* algorithms
Assumptions

Pressing

≤3 meters

Success

win the ball within 5s

Team A

https://en.wikipedia.org/wiki/Penalty_area#/media/File:Soccer_field_-_empty.svg

Team B
Approach

Machine Learning:

\[ \mathbf{X} = \text{Pressing start moment} \]
\[ \mathbf{Y} = \text{Successful or Unsuccessful} \]

1) \[ f(\mathbf{X}) = \mathbf{Y} \]

2) \[ g(\mathbf{X}) \sim f(\mathbf{X}) \]
Pattern Extraction

1) \[ g(x_1, x_2, x_3, \ldots, x_N) = \text{Prediction} \]

Increase/decrease

2) \[ g(x_1, x_2, x_3, \ldots, x_N) = \text{Prediction} \]

Increase/decrease
Attributes

\[ X = \text{Pressing start moment} \]

https://en.wikipedia.org/wiki/Penalty_area#/media/File:Soccer_field_-_empty.svg
### Accuracy

<table>
<thead>
<tr>
<th>Per Team: 1. Train/Test</th>
<th>2. Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>• First 80% matches</td>
<td>• Logistic Regression</td>
</tr>
<tr>
<td>• Last 20% matches</td>
<td>• Decision Tree</td>
</tr>
<tr>
<td></td>
<td>• Random Forest</td>
</tr>
<tr>
<td></td>
<td>• Naïve Bayes</td>
</tr>
<tr>
<td></td>
<td>• Light Gradient Boosting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lyon: 1. Accuracy</th>
<th>2. Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Training: 81%</td>
<td>• Majority Class: 55%</td>
</tr>
<tr>
<td>• Test: 71%</td>
<td></td>
</tr>
</tbody>
</table>
How to prepare for a new match?

https://www.oddsfan.com/board/match-preview-psg-vs-lyon/
Target zones to put Lyon under pressure

Number of Matches: 17
Pressure Success Rate Against Lyon: 57.78%

Players who lost the ball under pressure
Thank you for your attention

h.sotudeh@tue.nl

https://www.linkedin.com/in/hadisotudeh