Actions Speak Louder than Goals: Valuing Player Actions in Soccer

Tom Decroos, Lotte Bransen, Jan Van Haaren, Jesse Davis
KDD 2019
August 7, 2019
Key soccer analytics task: Valuing on-the-ball actions

Problem: Existing soccer stats value only a single type of action

Our contribution: A framework that values ALL on-the-ball actions
Challenge 1: Real-world action sequences are messy

- Missing or unrecorded actions
- Useless events (Weather change)
- Vendor-specific terminology
- Optional extra information

Pass

Tackle

Assist? High pass?
Contribution 1: SPADL is a unified and simple language for describing on-the-ball player actions

<table>
<thead>
<tr>
<th>Type:</th>
<th>Pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Player:</td>
<td>Eden Hazard</td>
</tr>
<tr>
<td>Team:</td>
<td>Chelsea</td>
</tr>
<tr>
<td>Result:</td>
<td>Success</td>
</tr>
<tr>
<td>Bodypart:</td>
<td>Foot</td>
</tr>
<tr>
<td>Time:</td>
<td>12min 36sec</td>
</tr>
<tr>
<td>Start:</td>
<td>x=53 y=15</td>
</tr>
<tr>
<td>End:</td>
<td>x=74 y=48</td>
</tr>
</tbody>
</table>

Converters available at: https://github.com/ML-KULeuven/socceraction/
Challenge 2: Most actions do not affect the score

A soccer game contains $\pm 1600$ actions
Most common final score: 1 – 0

What is the value of an action?
How good is a player?

Goal

High-level questions with no objective ground truth
Example action: Pass from Messi to Busquets
Action $a_i$ moves the game from state $S_{i-1}$ to state $S_i$
Contribution 2: The VAEP framework values an action by its expected impact on the score

**Intuition:** a good action $a_i$ for team $T$
(1) Increases the short-term probability of team $T$ scoring and/or
(2) Decreases the short-term probability of team $T$ conceding

VAEP value($a_i$) = $\Delta P_{scores}(a_i) - \Delta P_{concedes}(a_i)$

$\Delta P_{scores}(a_i) = P_{scores}(S_i, T) - P_{scores}(S_{i-1}, T)$
$\Delta P_{concedes}(a_i) = P_{concedes}(S_i, T) - P_{concedes}(S_{i-1}, T)$

Transformation from subjective task to objective ML task: estimating $P$
Our ML task: Estimate $P_{scores}(S_i, T)$ and $P_{concedes}(S_i, T)$

\[
\begin{align*}
X: & \text{ Features} \\
Y: & \text{ Labels} \\
F: & \text{ Probabilistic classifier}
\end{align*}
\]
$X$: Features that describe game state $S_i$

a) **Simple features**
- Action type
- Result
- ...

b) **Complex features**
- Distance to goal
- Time between actions
- ...

c) **Context features**
- Goal difference (e.g., +2, -1)
Y: Labels that capture $S_i$’s limited temporal influence

$$Y_{scores}(S_i, T) = \begin{cases} 
1 & \text{if team } T \text{ scores in the next 10 actions} \\
0 & \text{otherwise}
\end{cases}$$

$$Y_{concedes}(S_i, T) = \begin{cases} 
1 & \text{if team } T \text{ concedes in the next 10 actions} \\
0 & \text{otherwise}
\end{cases}$$
F: Probabilistic classifier

Brier score
- accuracy
- calibration

ROC AUC
works well for unbalanced data sets
Our ML task: Estimate $P_{\text{scores}}(S_i, T)$ and $P_{\text{concedes}}(S_i, T)$

$X$: Features

Simple features + Complex features + Context features

$Y$: Labels

1 if team T scores/concedes in the next 10 actions

$F$: Probabilistic classifier

CatBoost
Our soccer analytics task: **Value on-the-ball actions**

**X:** Features

- Simple features + Complex features + Context features

**Y:** Labels

- 1 if team T scores/concedes in the next 10 actions

**F:** Probabilistic classifier

- CatBoost

**Formula:**  

\[ \text{VAEP value}(a_i) = \Delta P_{scores}(a_i) - \Delta P_{concedes}(a_i) \]
Intuitive illustration of VAEP values:
Barcelona’s 3-0 goal vs Real Madrid (Dec 23, 2017)

Phase starts here
Applications in player scouting

Rating players
Identifying top players
Comparing playing styles
The big question
Our soccer data

7 European competitions
Premier League, La Liga, Eredivisie, ...

5 seasons
2012/13 – 2017/18

11,565 games
14,427,803 actions
Our soccer data

7 European competitions
Premier League, La Liga, Eredivisie, ...

5 seasons
2012/13 – 2017/18

11,565 games

14,427,803 actions

Test data
Rating players on expected score impact

Romelu Lukaku
Striker at Manchester United

- 2869 minutes
- 966 actions
- 16 goals
- 7 assists

Trent Alexander-Arnold
Defender at Liverpool

- 1575 minutes
- 1528 actions
- 1 goal
- 2 assists

Naive metric: goals + assists per 90 minutes

0.72 vs 0.11

Let’s rate players on ALL their actions instead
Rating players on expected score impact

\[ \frac{1}{n} \sum_{i}^{n} \text{VAEP Value}(a_i) \]

Average action value (quality)

Actions per 90 minutes (quantity)
Rating players on expected score impact

\[ = \frac{1}{n} \sum_{i}^{n} \text{VAEP Value}(a_i) \]

Average action value (quality)

VAEP rating 0.40
VAEP rating 0.29

= quantity * quality
Rating players on expected score impact

All players in Spain and England in 2017/18

Goalkeepers

Average action value (quality)

0.020

0.015

0.010

0.005

0.000

0

25

50

75

100

125

Actions per 90 minutes (quantity)
Top-5 players in the 2017/18 Premier League

<table>
<thead>
<tr>
<th>Rank</th>
<th>Player</th>
<th>VAEP rating</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Philippe Coutinho</td>
<td>0.90</td>
<td>€ 140m</td>
</tr>
<tr>
<td>2</td>
<td>Mohammed Salah</td>
<td>0.82</td>
<td>€ 150m</td>
</tr>
<tr>
<td>3</td>
<td>Kevin De Bruyne</td>
<td>0.64</td>
<td>€ 150m</td>
</tr>
<tr>
<td>4</td>
<td>Eden Hazard</td>
<td>0.64</td>
<td>€ 150m</td>
</tr>
<tr>
<td>5</td>
<td>Riyad Mahrez</td>
<td>0.63</td>
<td>€ 60m</td>
</tr>
</tbody>
</table>
# Top-5 U21 players in the 2017/18 Dutch League

<table>
<thead>
<tr>
<th>Rank</th>
<th>Player</th>
<th>Team</th>
<th>Age</th>
<th>VAEP rating</th>
<th>June 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>David Neres</td>
<td>Ajax</td>
<td>21</td>
<td>0.62</td>
<td>€ 20m</td>
</tr>
<tr>
<td>2</td>
<td>Mason Mount</td>
<td>Vitesse</td>
<td>19</td>
<td>0.62</td>
<td>€ 4m</td>
</tr>
<tr>
<td>3</td>
<td>Frenkie de Jong</td>
<td>Ajax</td>
<td>20</td>
<td>0.50</td>
<td>€ 7m</td>
</tr>
<tr>
<td>4</td>
<td>Steven Bergwijn</td>
<td>PSV</td>
<td>20</td>
<td>0.49</td>
<td>€ 12m</td>
</tr>
<tr>
<td>5</td>
<td>Donny van de Beek</td>
<td>Ajax</td>
<td>21</td>
<td>0.47</td>
<td>€ 14m</td>
</tr>
</tbody>
</table>
# Top-5 U21 players in the 2017/18 Dutch League

<table>
<thead>
<tr>
<th>Rank</th>
<th>Player</th>
<th>Team</th>
<th>Age</th>
<th>VAEP rating</th>
<th>June 2018</th>
<th>June 2019</th>
<th>Price delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>David Neres</td>
<td>Ajax</td>
<td>21</td>
<td>0.62</td>
<td>€ 20m</td>
<td>€ 45m</td>
<td>+ €25m</td>
</tr>
<tr>
<td>2</td>
<td>Mason Mount</td>
<td>Vitesse</td>
<td>19</td>
<td>0.62</td>
<td>€ 4m</td>
<td>€ 12m</td>
<td>+ €8m</td>
</tr>
<tr>
<td>3</td>
<td>Frenkie de Jong</td>
<td>Ajax</td>
<td>20</td>
<td>0.50</td>
<td>€ 7m</td>
<td>€ 85m</td>
<td>+ €78m</td>
</tr>
<tr>
<td>4</td>
<td>Steven Bergwijn</td>
<td>PSV</td>
<td>20</td>
<td>0.49</td>
<td>€ 12m</td>
<td>€ 35m</td>
<td>+ €23m</td>
</tr>
<tr>
<td>5</td>
<td>Donny van de Beek</td>
<td>Ajax</td>
<td>21</td>
<td>0.47</td>
<td>€ 14m</td>
<td>€40m</td>
<td>+ €26m</td>
</tr>
</tbody>
</table>
Can **Hazard** replace **Ronaldo**?

<table>
<thead>
<tr>
<th>Action</th>
<th>VAEP rating 0.64</th>
<th>VAEP rating 0.61</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Shots</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>Dribbles</td>
<td>5</td>
</tr>
<tr>
<td>47</td>
<td>Passes</td>
<td><strong>27</strong></td>
</tr>
<tr>
<td>2</td>
<td>Crosses</td>
<td>1</td>
</tr>
</tbody>
</table>

**Actions per 90 minutes**
Can **Hazard** replace **Ronaldo**?

**VAEP rating 0.64**
- 2 Shots
- 11 Dribbles
- 47 Passes
- 2 Crosses

**VAEP rating 0.61**
- 5 Shots
- 5 Dribbles
- 27 Passes
- 1 Crosses

Total value per 90 minutes
The big question
Who is better, Ronaldo or Messi?

All players in Spain and England in 2017/18

Average action value (quality)

Goalkeepers

Attackers

Midfielders

Defenders

Actions per 90 minutes (quantity)
Who is better, Ronaldo or Messi?

All players in Spain and England in 2017/18 except Ronaldo and Messi.

- Attackers
- Midfielders
- Defenders
- Goalkeepers

Average action value (quality) vs. Actions per 90 minutes (quantity).
Who is better, Ronaldo or Messi?
Who is better, Ronaldo or Messi?
Online resources

https://github.com/ML-KULeuven/socceraction/
- pip install socceraction
- Example notebooks demonstrating the full pipeline with free StatsBomb data

https://www.scisports.com/services/insight/
Concluding thoughts

Challenges:
- Real-world soccer data != UCI data sets
- Often no ground truth available

Valuing all on-the-ball player actions:
- Captures information ignored by existing soccer stats
- Has many use cases, e.g., player scouting

Messi > Ronaldo 😊
Contributions

1. **SPADL**: a unified and simple language for soccer actions

2. **VAEP**: a framework to assign values to ALL actions in soccer

3. Use cases relevant for scouting