Football Analytics

Hadi Sotudeh

January 2020

Agenda

- Introduction
- Datasets
- Applications
- Case Study 1
- Case Study 2
- Q & A

Introduction

https://images.app.goo.gl/PcA6neMBwpYwQ1aw8

Datasets

Event Data:

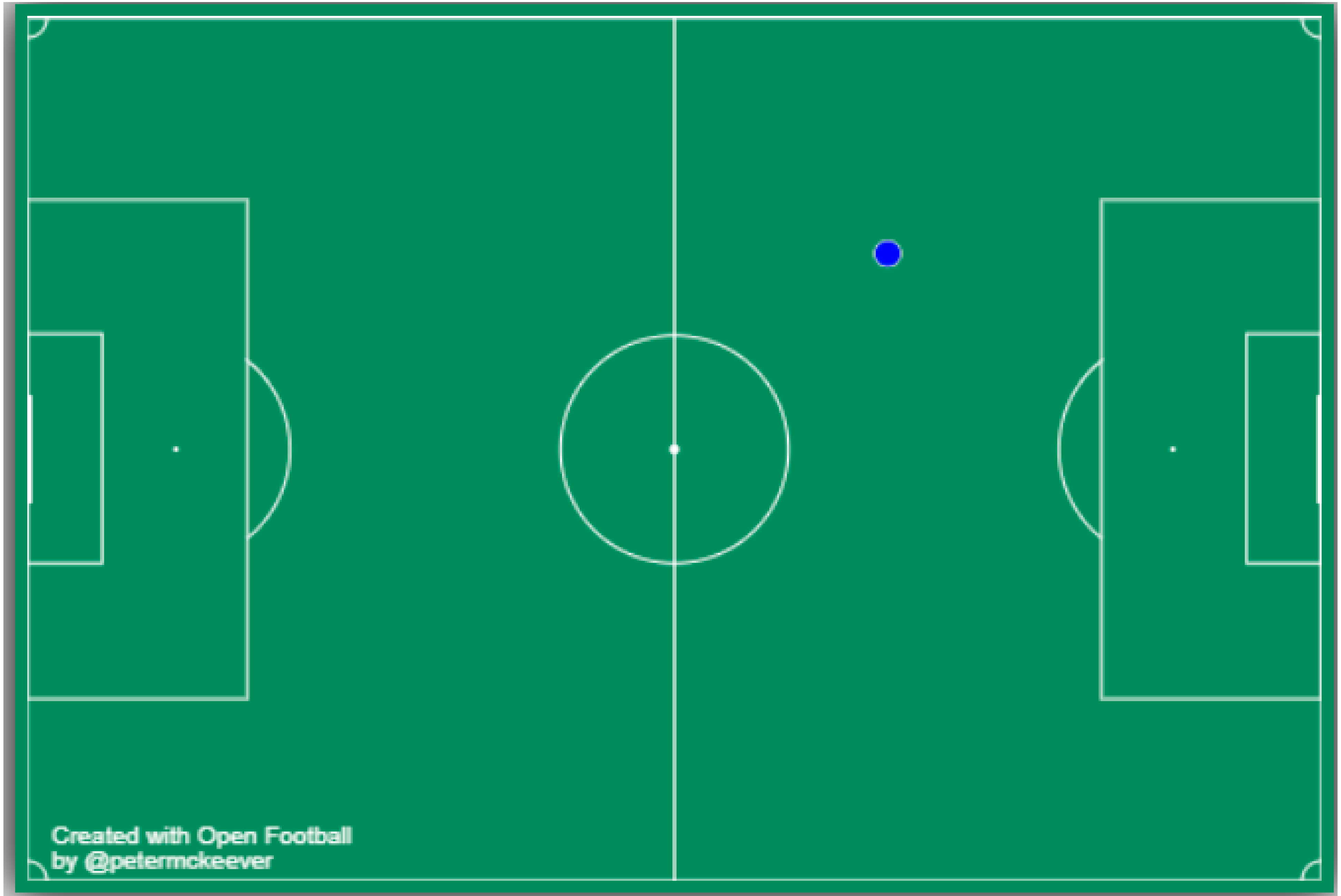
- On-ball actions
- ~2000 events per match
- Easy to query:
  - All the final third passes by Karimi
- Without context

Opta - Data Collection

https://www.youtube.com/watch?v=Z52mmpF2wLg

Opta - Data Collection

https://www.youtube.com/watch?v=Z52mmpF2wLg



(69.8,50.71999969482422)

http://openfootball.club


Opta - Data Collection
https://www.youtube.com/watch?v=Z52mmpF2wLg

Event Data Example:


```xml
<?xml version="1.0" encoding="UTF-8"?>
<!-- Copyright 2001-2017 Opta Sportsdata Ltd. All rights reserved. -->

<!-- PRODUCTION HEADER
     produced on:        valde-jobq-a02.nexus.opta.net
     production time:    20170119T183125,93Z
     production module:  Opta::Feed::XML::Soccer::F24
-->
<Games timestamp="2017-01-19T18:31:24">
  <Game id="853308" away_team_id="1028" away_team_name="Caen" competition_id="24" competition_name="French Ligue 1" game_date="2017-01-18T19:00:00" home_team_id
    <Event id="1670359607" event_id="1" type_id="34" period_id="16" min="0" sec="0" team_id="1028" outcome="1" x="0.0" y="0.0" timestamp="2017-01-18T18:08:25.86
      <Q id="1549473786" qualifier_id="59" value="1, 24, 11, 18, 5, 28, 7, 6, 26, 25, 12, 2, 4, 10, 17, 20, 22, 40" />
      <Q id="2104355445" qualifier_id="131" value="1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 0, 0, 0, 0, 0, 0, 0" />
      <Q id="1042689974" qualifier_id="30" value="9997, 203368, 42761, 119691, 12845, 44287, 184095, 86281, 60319, 45082, 50996, 7103, 97443, 6596, 199666, 5861
      <Q id="1458083580" qualifier_id="44" value="1, 2, 2, 2, 2, 2, 3, 3, 4, 3, 3, 5, 5, 5, 5, 5, 5, 5" />
      <Q id="1254524248" qualifier_id="227" value="0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0" />
      <Q id="697029712" qualifier_id="130" value="11" />
      <Q id="230429698" qualifier_id="194" value="45082" />
    </Event>
    <Event id="802446034" event_id="2" type_id="34" period_id="16" min="0" sec="0" team_id="430" outcome="1" x="0.0" y="0.0" timestamp="2017-01-18T18:08:28.471"
      <Q id="1483710138" qualifier_id="197" value="654" />
      <Q id="733972045" qualifier_id="131" value="1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 0, 0, 0, 0, 0, 0, 0" />
      <Q id="1873267406" qualifier_id="44" value="1, 2, 2, 3, 2, 2, 3, 3, 4, 4, 3, 5, 5, 5, 5, 5, 5, 5" />
      <Q id="1305768225" qualifier_id="130" value="2" />
      <Q id="1130563308" qualifier_id="59" value="30, 15, 6, 27, 26, 3, 7, 28, 10, 9, 14, 2, 4, 8, 13, 18, 19, 40" />
      <Q id="1820025184" qualifier_id="30" value="97288, 183779, 116536, 42586, 102766, 165659, 119735, 193539, 193409, 103124, 229608, 148616, 33946, 115961, 4
      <Q id="875720895" qualifier_id="227" value="0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0" />
      <Q id="585487239" qualifier_id="194" value="42586" />
    </Event>
    <Event id="289241211" event_id="2" type_id="32" period_id="1" min="0" sec="0" team_id="1028" outcome="1" x="0.0" y="0.0" timestamp="2017-01-18T19:00:08.677"
      <Q id="1553063144" qualifier_id="127" value="Right to Left" />
    </Event>
    <Event id="565328556" event_id="3" type_id="32" period_id="1" min="0" sec="0" team_id="430" outcome="1" x="0.0" y="0.0" timestamp="2017-01-18T19:00:08.984"
      <Q id="944425465" qualifier_id="127" value="Left to Right" />
    </Event>
    <Event id="963150344" event_id="4" type_id="1" period_id="1" min="0" sec="0" player_id="193409" team_id="430" outcome="1" x="49.5" y="50.6" timestamp="2017-
      <Q id="1024498704" qualifier_id="140" value="34.2" />
      <Q id="1483160354" qualifier_id="212" value="17.4" />
      <Q id="1687182276" qualifier_id="279" value="S" />
```

Assignment 1

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

tinyurl.com/ty6op7q

https://images.app.goo.gl/mM3o5nA2VVnRWwrv6

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

https://images.app.goo.gl/zPm3298W6PcxCCpC6

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)

https://en.wikipedia.org/wiki/SportVU

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¶

http://pjreddie.com/yolo/

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.

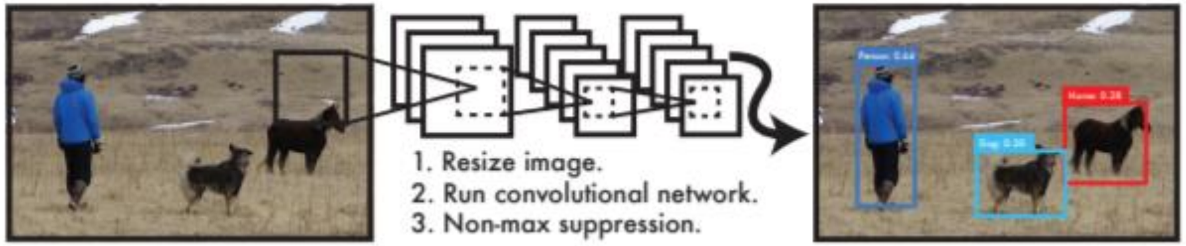
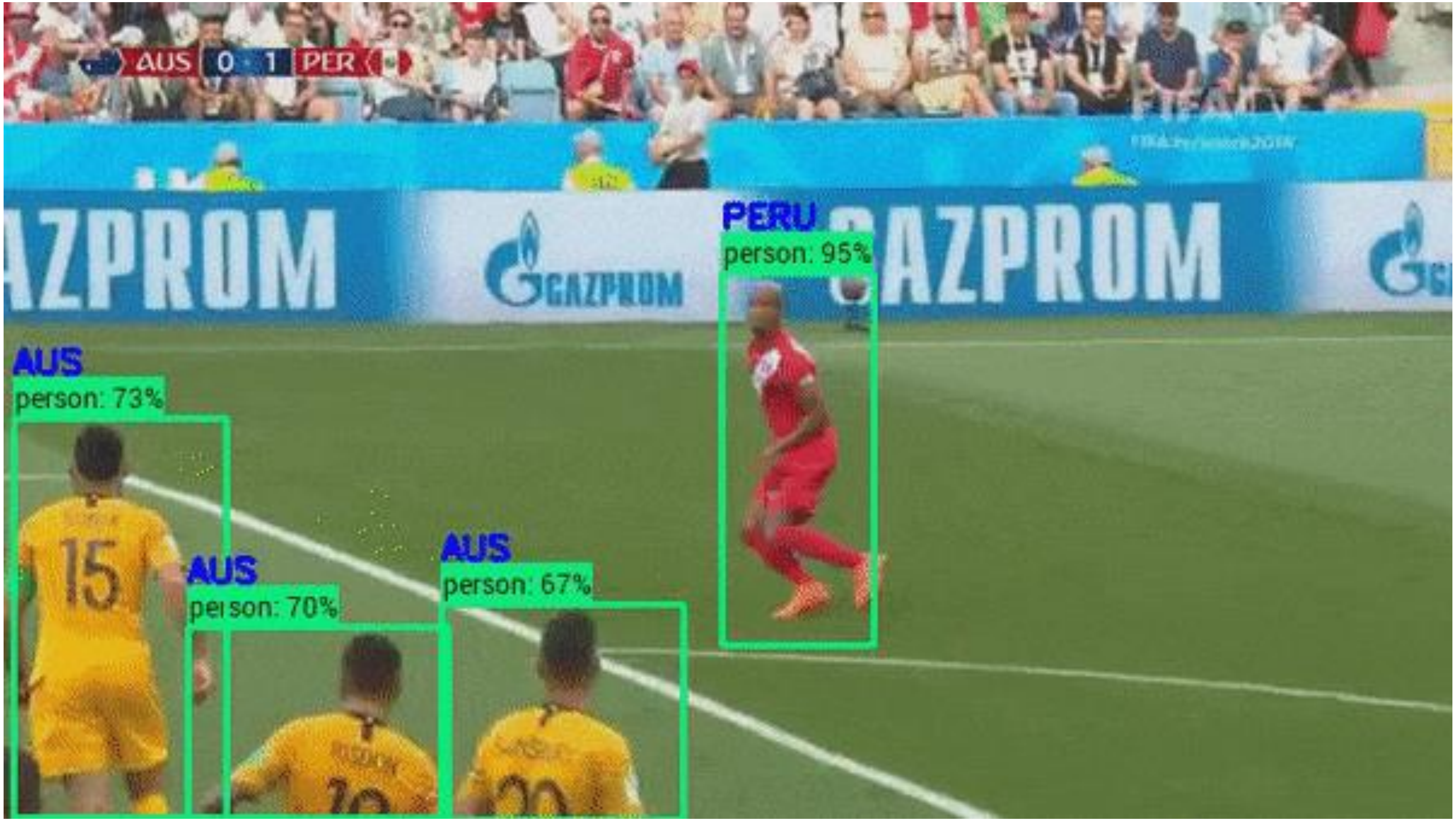
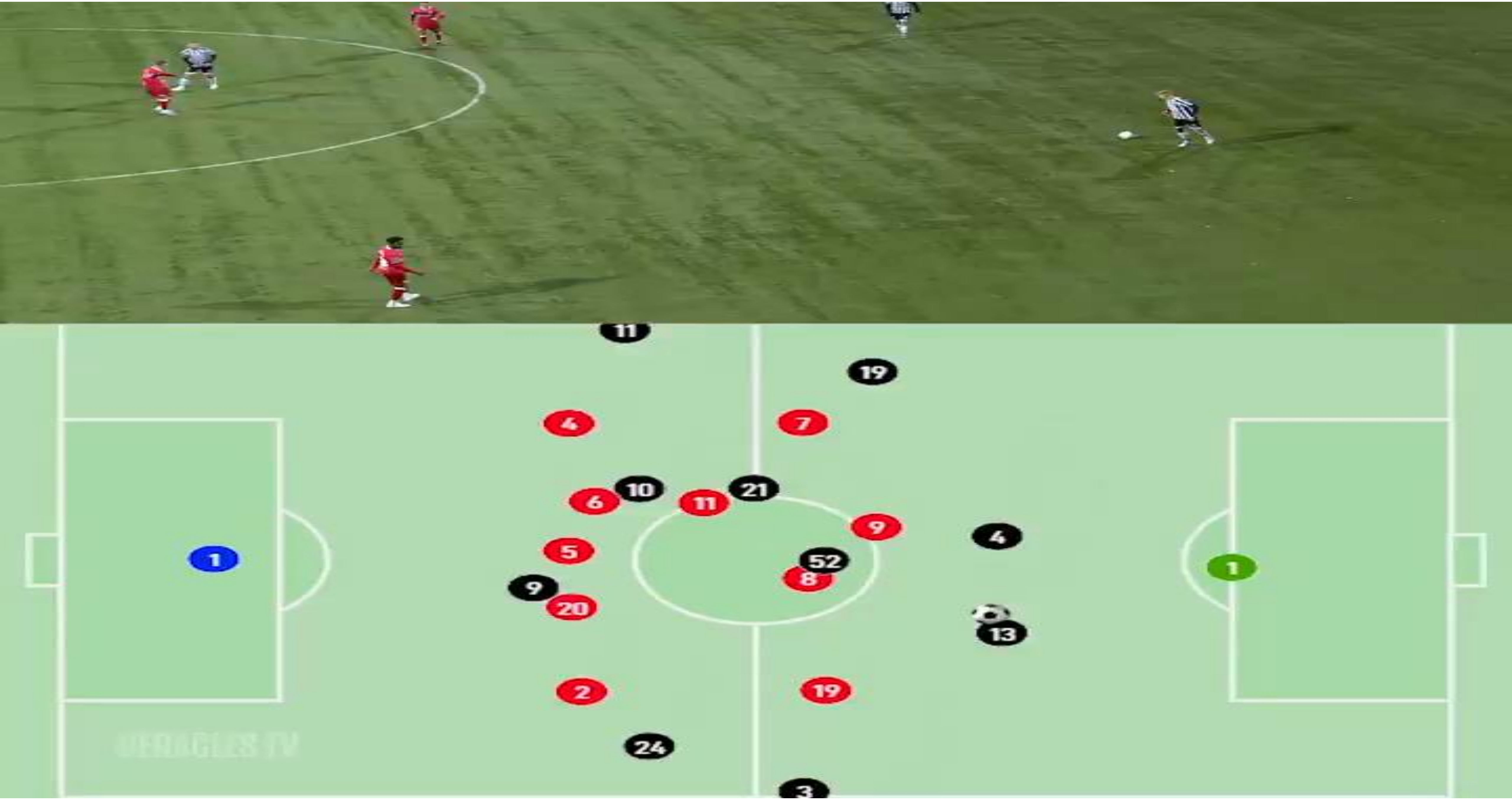


Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.

https://arxiv.org/pdf/1506.02640v5.pdf



https://www.kdnuggets.com/2018/07/analyze-soccer-game-using-tensorflow-object-detection-opencv.html



https://twitter.com/HeraclesAlmelo/status/1194254478842769409

Tracking Data Example

PLAYER_ID	TEAM_FIXTURE	HALF	PLAYER_X_POSITION	PLAYER_Y_POSITION	TIME_VALUE	X_POSITION	Y_POSITION
723245	Lyon	First Half	52.64	33.82	0.0	52.5	34.0
347689	Paris SG	First Half	42.41	33.71	0.0	52.5	34.0
346567	Paris SG	First Half	41.90	57.61	0.0	52.5	34.0
347763	Paris SG	First Half	49.40	52.81	0.0	52.5	34.0
348502	Paris SG	First Half	38.18	38.07	0.0	52.5	34.0
347858	Lyon	First Half	71.99	14.62	0.0	52.5	34.0
464719	Lyon	First Half	52.91	10.39	0.0	52.5	34.0
375843	Paris SG	First Half	33.51	36.63	0.0	52.5	34.0
437146	Paris SG	First Half	52.35	43.78	0.0	52.5	34.0
440664	Lyon	First Half	74.17	42.43	0.0	52.5	34.0

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.”

https://images.app.goo.gl/Gc3uTGYJFZq6yvbDA

Actions Speak Louder than Goals: Valuing Player Actions in Soccer

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

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

Jan Van Haaren
SciSports
Amersfoort, Netherlands
j.vanhaaren@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

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

https://arxiv.org/pdf/1802.07127.pdf

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*).

https://arxiv.org/pdf/1802.07127.pdf

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

$$
\begin{aligned}
P_{scores}(S_i, h) &= P(goal(h) \in F_i^k | S_i) \\
P_{concedes}(S_i, h) &= P(goal(v) \in F_i^k | S_i) \\
P_{scores}(S_i, v) &= P(goal(v) \in F_i^k | S_i) \\
P_{concedes}(S_i, v) &= P(goal(h) \in F_i^k | S_i)
\end{aligned}
$$

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

https://arxiv.org/pdf/1802.07127.pdf

Binary classification:

- Any machine learning algorithm that predicts a probability:
  - Logistic Regression, Random Forest, or Neural Network
- The probability estimates should be well-calibrated
  - https://scikit-learn.org/stable/modules/calibration.html

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

https://arxiv.org/pdf/1802.07127.pdf

$$\Delta P_{scores}(a_i, x) = P_{scores}(S_i, x) - P_{scores}(S_{i-1}, x). \qquad (1)$$

$$\Delta P_{concedes}(a_i, x) = P_{concedes}(S_i, x) - P_{concedes}(S_{i-1}, x). \qquad (2)$$

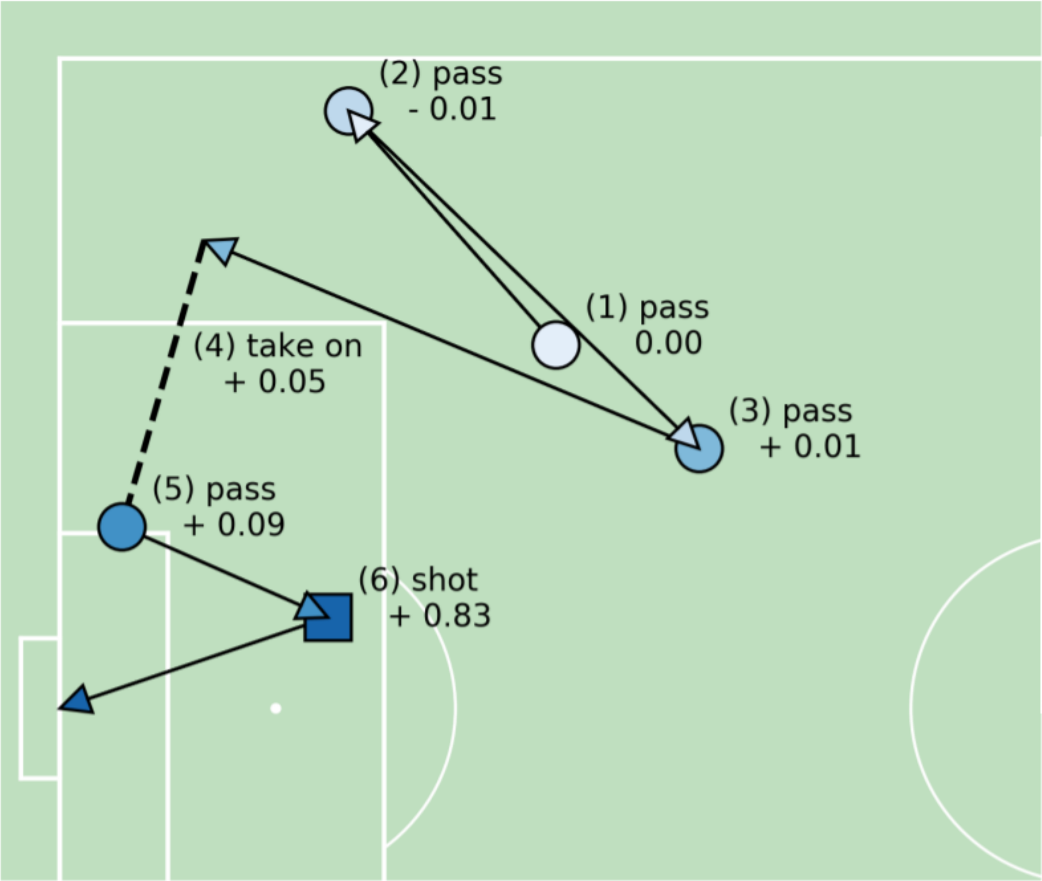
https://arxiv.org/pdf/1802.07127.pdf

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

https://arxiv.org/pdf/1802.07127.pdf

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



	TIME	PLAYER	ACTION	P_{scores}	VALUE
1	92m4s	S. Busquets	pass	0.03	0.00
2	92m6s	L. Messi	pass	0.02	- 0.01
3	92m8s	S. Busquets	pass	0.03	+ 0.01
4	92m11s	L. Messi	take on	0.08	+ 0.05
5	92m12s	L. Messi	pass	0.17	+ 0.09
6	92m14s	A. Vidal	shot	1.00	+ 0.83

https://arxiv.org/pdf/1802.07127.pdf

VAEP (Valuing Actions by Estimating Probabilities)

$$V(a_i, x) = \Delta P_{scores}(a_i, x) + (-\Delta P_{concedes}(a_i, x)) \qquad (3)$$

$$rating(p) = \frac{90}{m} \sum_{a_i \in A_p^T} V(a_i), \qquad (4)$$

https://arxiv.org/pdf/1802.07127.pdf

Results

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

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

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

https://arxiv.org/pdf/1802.07127.pdf

Results

Born after 1/1/1997

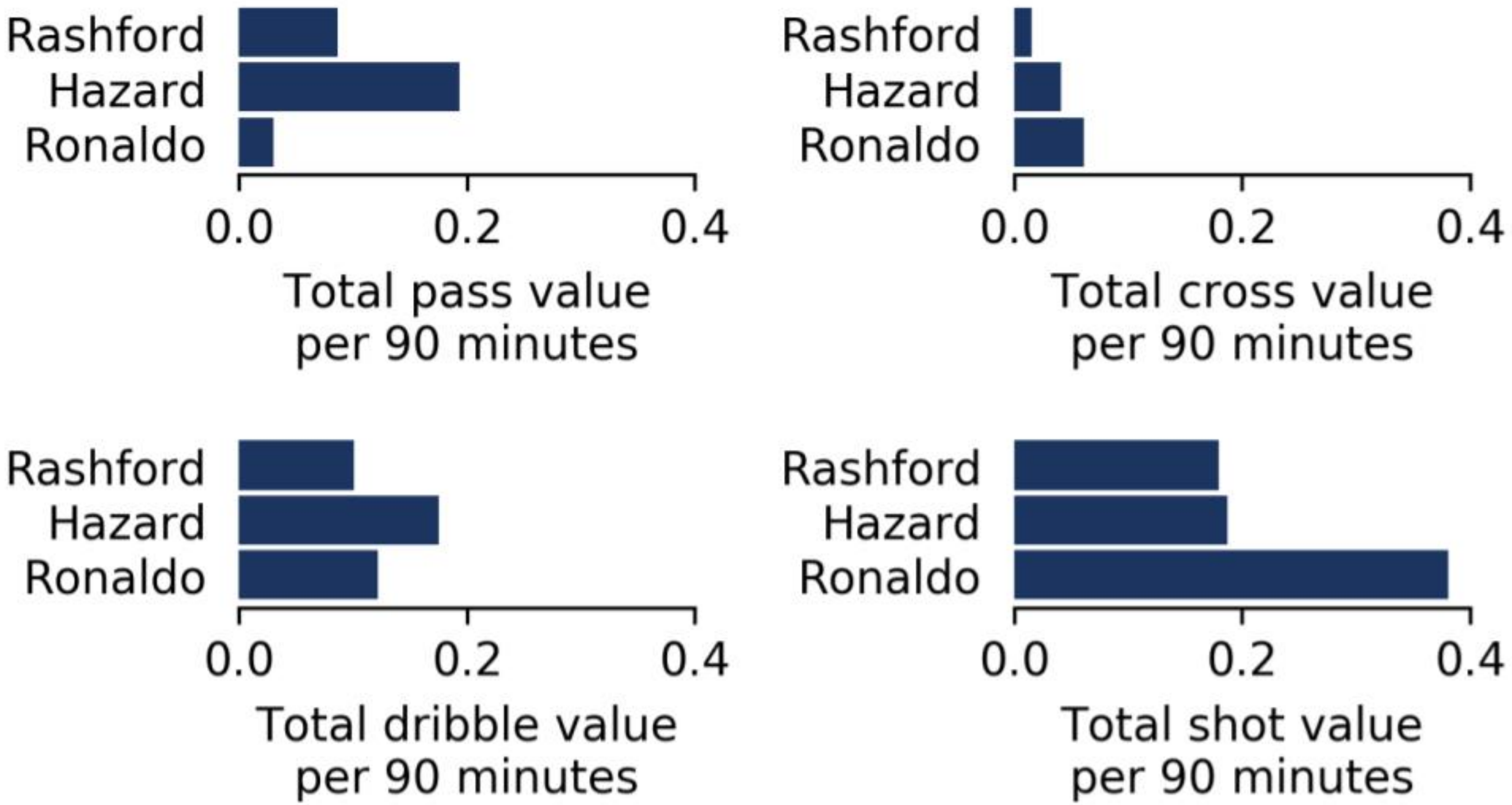
2017/2018 season

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

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

https://arxiv.org/pdf/1802.07127.pdf

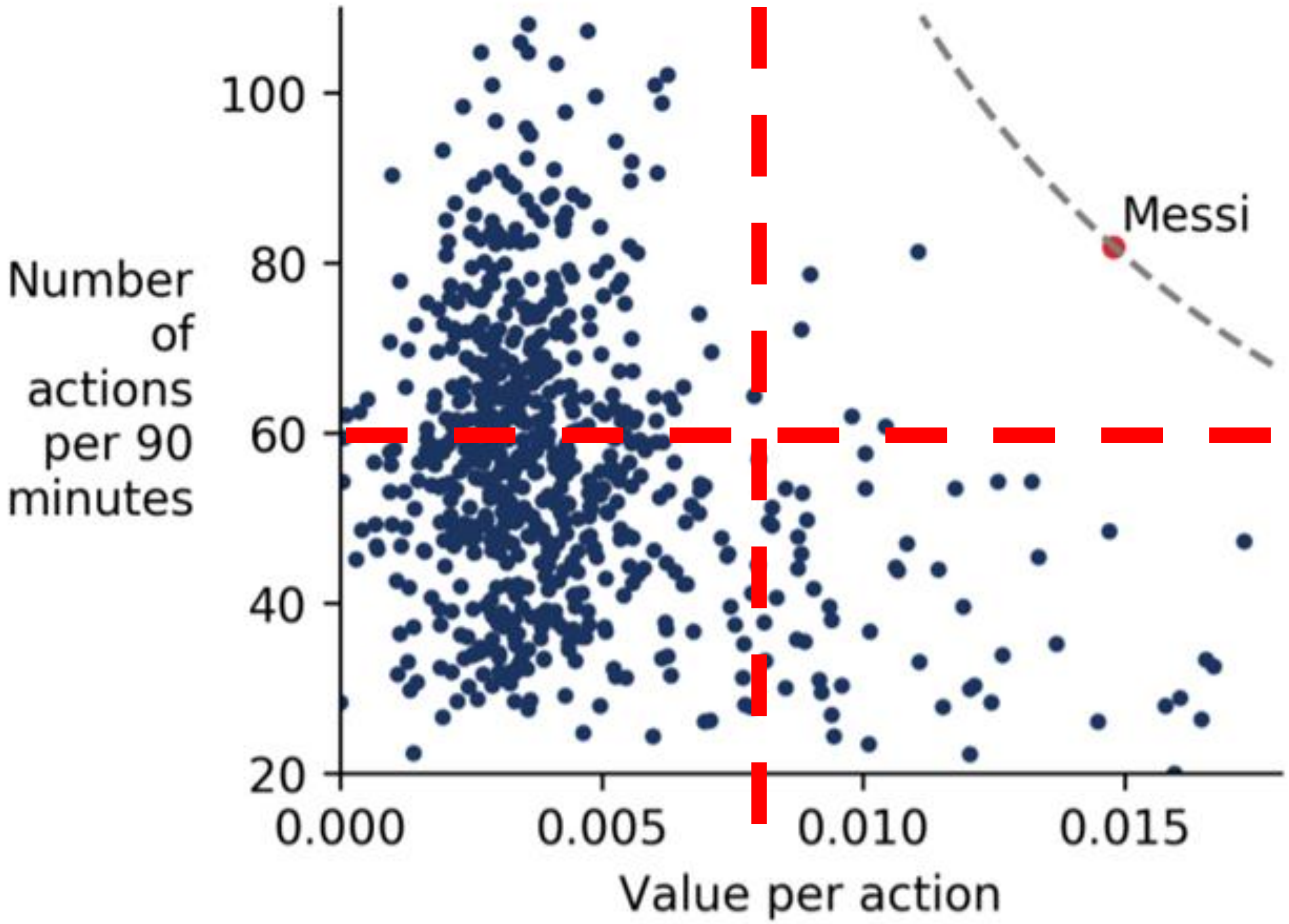
Results



(b) Playing style of replacement players for Ronaldo

https://arxiv.org/pdf/1802.07127.pdf

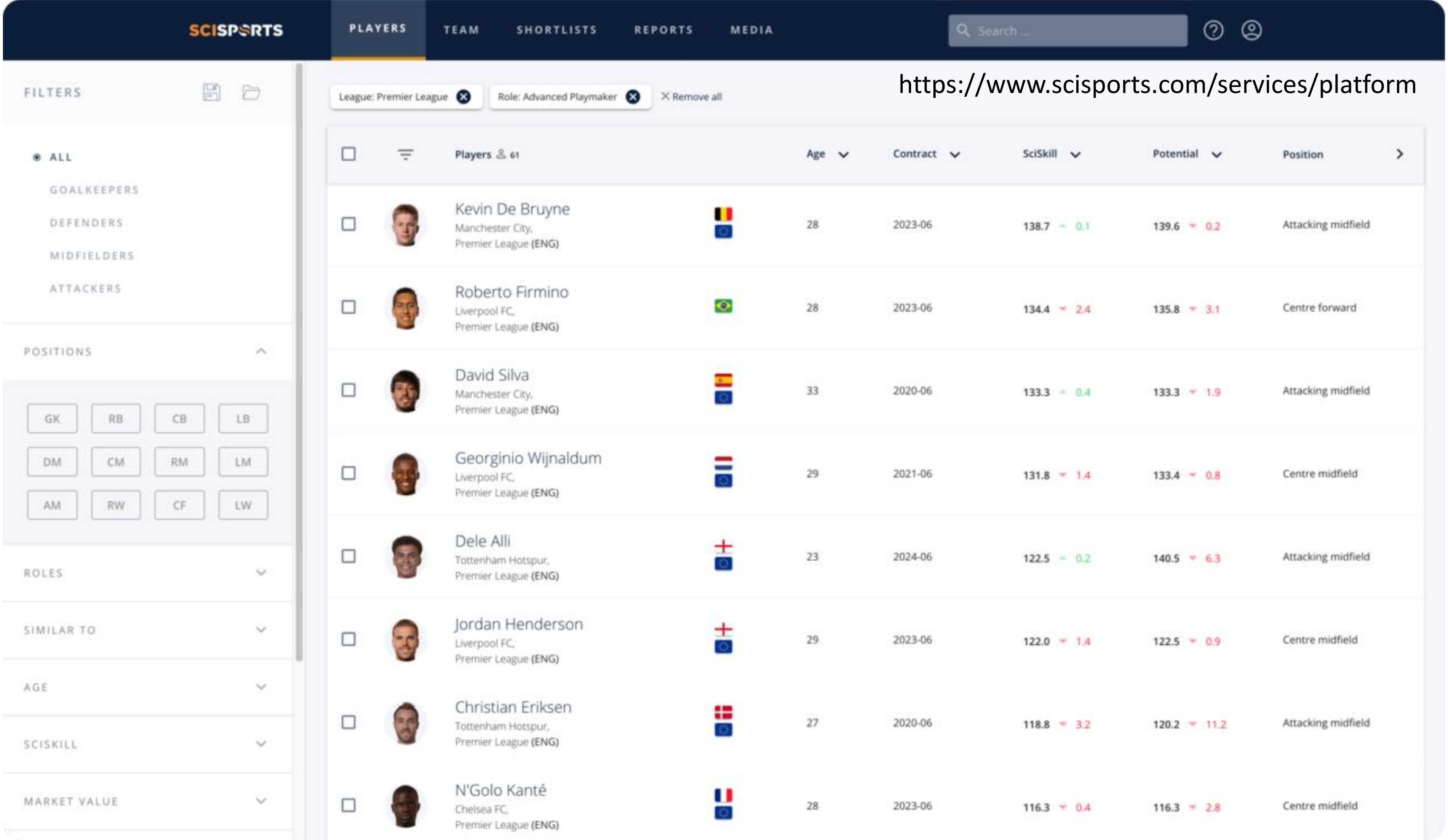
Players in the 2017/2018 season who played at least 900 mins



(a) All players

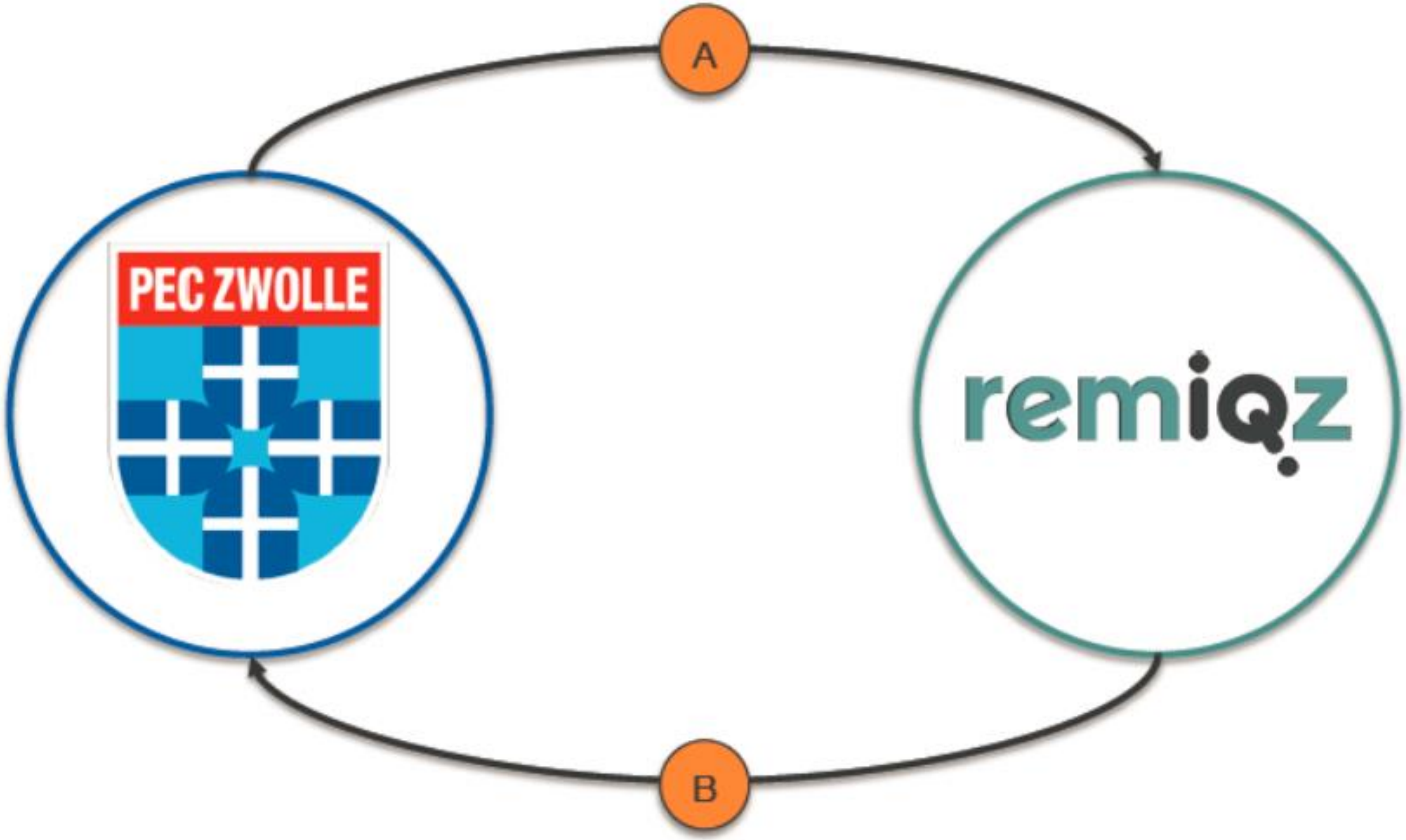
https://arxiv.org/pdf/1802.07127.pdf

https://www.scisports.com/services/platform



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

birthArea	birthDate	currentNationalTeamId	currentTeamId	firstName	foot	height	lastName	passportArea	role	shortName	weight	wyId
Netherlands	7/25/1998	671	23568	Thijmen	right	196	Nijhuis	Netherlands	Goalkeeper	T. Nijhuis	83	365439
Netherlands	8/12/1999	664	9	Matthijs	right	188	de Ligt	Netherlands	Defender	M. de Ligt	89	365443
Netherlands	2/28/1998	670	10	Teun	left	183	Koopmeiners	Netherlands	Midfielder	T. Koopmeiner	71	365444
Netherlands	3/13/1998	670	32	Jay-Roy	right	193	Grot	Suriname	Forward	J. Grot	98	365445
Netherlands	1/19/1999	670	11	Donyell	right	179	Malen	Suriname	Forward	D. Malen	78	365446
Austria	2/4/1998	9116	680	Maximilian	left	189	WÃ¶ber	Austria	Defender	M. WÃ¶ber	82	367503
England	1/26/1998	2421	39826	Easah	left	188	Suliman	England	Defender	E. Suliman	78	368350

KPIs.csv

wyId	accelerations	aerialDuels	aerialDuelsWon	assists	backPasses	corners	crosses	directRedCard	dribbles	forwardPasses	fouls	goals	headShots	interceptions	longPasses	losses	matches
149157	1	19	5	0	102	0	7	0	19	175	12	0	0	53	34	97	11
218554	0	0	0	0	0	0	0	0	0	10	0	0	0	1	5	2	1
216568	15	11	1	2	85	5	28	0	70	90	12	3	1	25	19	103	18
221036	1	17	10	0	30	5	1	0	5	80	9	0	0	40	35	59	9
221037	3	7	2	0	35	0	8	0	26	59	3	0	0	31	12	43	6

tinyurl.com/u7ybrqg

Position: Defenders

Main position: Right-Back

Age group: All

Year: All

Nationality: Netherlands

Confederation: Worldwide

Show

Compact | Detailed | Gallery

#	Player	Age	Nat.	Club	Market value
1	Hans Hateboer Right-Back	25			€15.00m
2	Denzel Dumfries Right-Back	23			€13.50m
3	Joël Veltman Right-Back	27			€10.00m
4	Timothy Fosu-Mensah Right-Back	21			€7.00m
5	Kenny Tete Right-Back	24			€7.00m
6	Rick Karsdorp Right-Back	24			€7.00m

https://www.transfermarkt.com

External Datasets

https://sofifa.com

M. Taremi (ID: 241788) FIFA 20 DEC 18, 2019

Mehdi Taremi CF LW Age 26 (Jul 18, 1992) 6'2" 176lbs

74 Overall Rating | 74 Potential | 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
- 56 Crossing
- 74 Finishing
- 69 Heading Accuracy
- 68 Short Passing
- 71 Volleys

Skill
- 77 Dribbling
- 71 Curve
- 43 FK Accuracy
- 47 Long Passing
- 78 Ball Control

Movement
- 68 Acceleration
- 67 Sprint Speed
- 67 Agility
- 74 Reactions
- 60 Balance

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

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

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

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

Talent detection in soccer using a one-class support vector machine

Jauhiainen Susanne[1]*, Forsman Hannele[2], Äyrämö Sami[1], and Kauppi Jukka-Pekka[1]

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

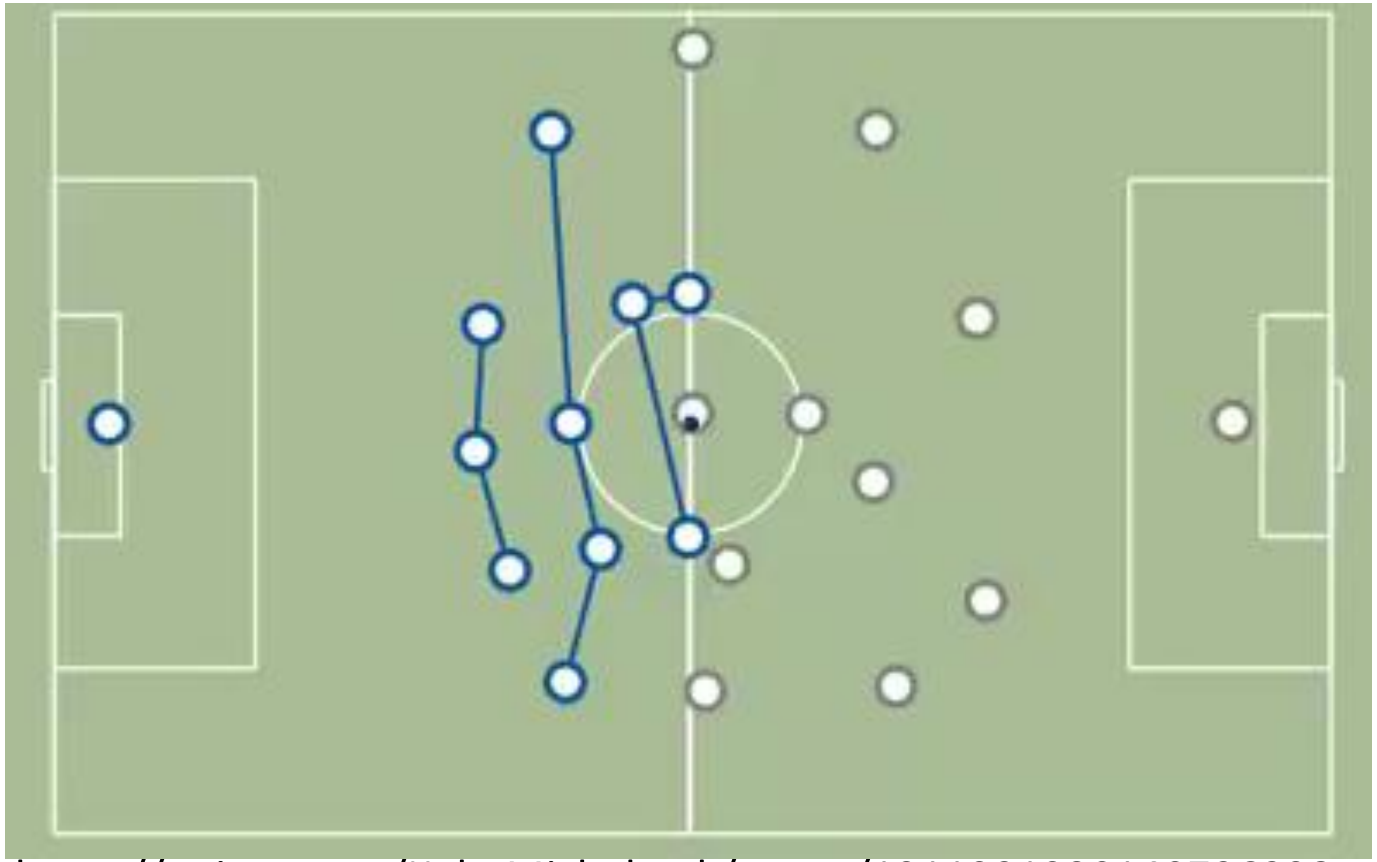
Abstract — A large variety of different features are needed to succeed in soccer and therefore talent identification is a very complex process. In this study, unsupervised anomaly detection was used to get new insight into talent identification. The aim was to build an automatic "talent detector" from 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.

https://content.sciendo.com/view/journals/ijcss/18/3/article-p125.xml

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


TW
تلوبیون
مستند " منچسترسیتی ، همه یا هیچ

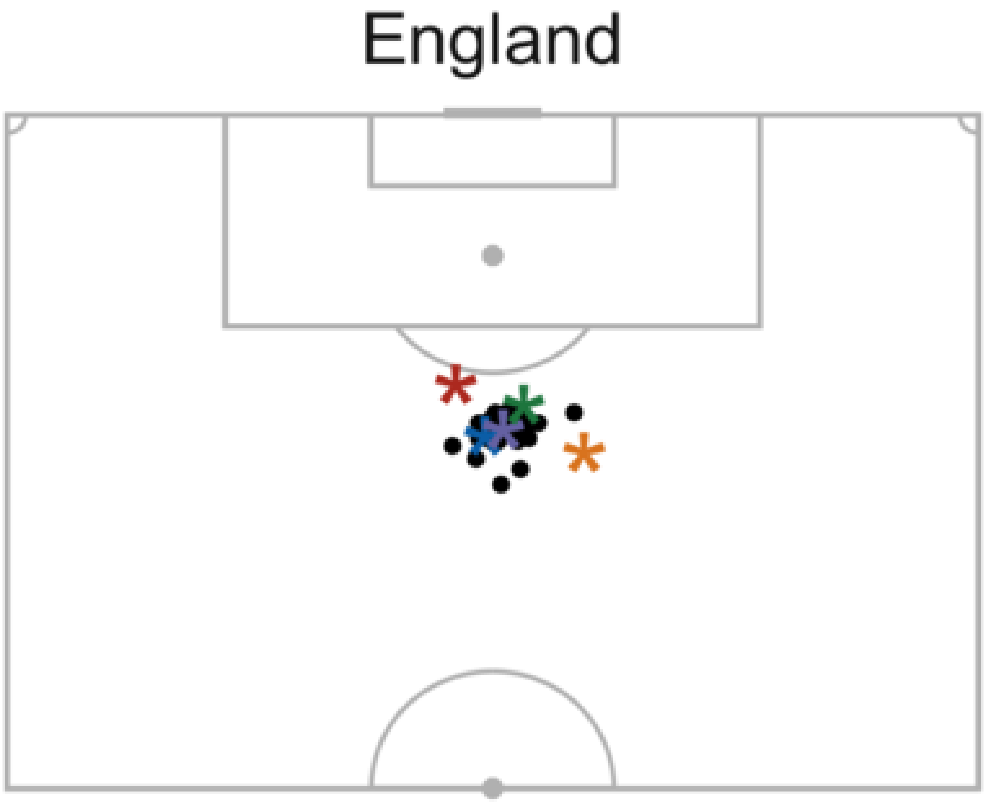
Coding Data:

https://images.app.goo.gl/MovFXrnSDbcvyUKj9

Half-Time Break!

Challenges

1. Having a high data quality:



Data: * A * B * C * D * Mean Google Survey

https://www.americansocceranalysis.com/home/2019/9/30/shots-in-the-dark-how-data-providers-tell-us-different-versions-of-what-happened

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

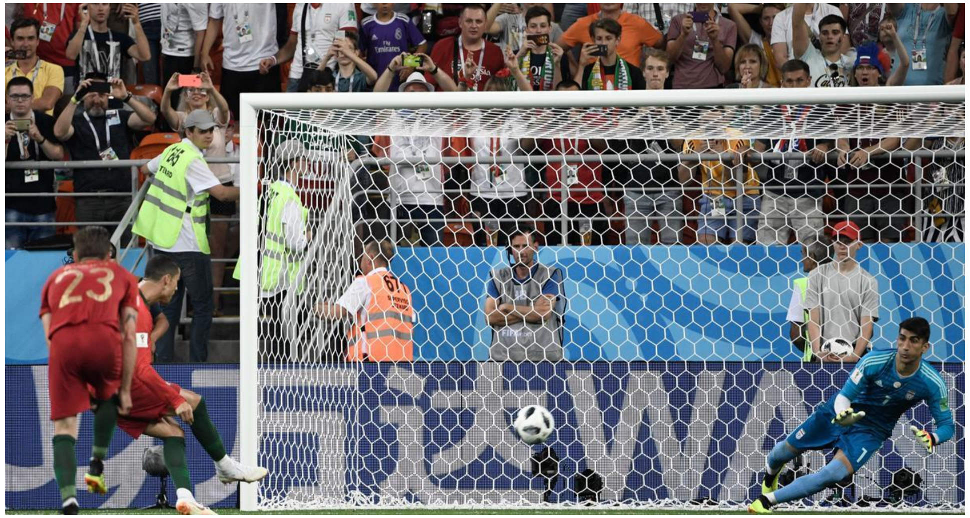
How does a football team possess the ball?

https://theprideoflondon.com/2018/12/12/chelsea-maurizio-sarri-eden-hazard-false-nine

Event Data:

FIFA WORLD CUP
RUSSIA 2018

Iran - Portugal



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


```
"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"
  }
```

Ball possession process

caseID	action	type	play_pattern	recipient	startTime	period	duration	endTime
7557160	Pass	Free Kick	From Free Kick	Omid Ebrahimi	00:48:41.660	2	1.6	00:48:43.260000
7557160	Pass	None	From Free Kick	Sardar Azmoun	00:48:46.460	2	2.76	00:48:49.220000
7557160	Pass	None	From Free Kick	Saman Ghoddos	00:48:49.220	2	1.12	00:48:50.340000
7557160	Shot	Open Play	From Free Kick		00:48:51.753	2	0.107	00:48:51.860000
7557160	Block	None	From Free Kick		00:48:51.860	2		None
7557160	Goal Keeper	Shot Faced	From Free Kick		00:48:51.995	2		None
7557160	Ball Recovery	None	From Free Kick		00:48:53.380	2		None
7557160	Shot	Open Play	From Free Kick		00:48:53.993	2	0.44	00:48:54.433000
7557160	Goal Keeper	Shot Faced	From Free Kick		00:48:54.433	2		None

possession_team	team_action	player	body_part	start_X	start_Y	end_X	end_Y	result
Iran	Iran	Ramin Rezaeian	Right Foot	28.0	78.0	28.0	54.0	None
Iran	Iran	Omid Ebrahimi	Right Foot	35.0	51.0	86.0	53.0	None
Iran	Iran	Sardar Azmoun		84.0	40.0	97.0	39.0	None
Iran	Iran	Saman Ghoddos	Right Foot	97.0	45.0	101.0	45.0	Blocked
Iran	Portugal	JosÃ© Miguel da Rocha Fonte		20.0	36.0			None
Iran	Portugal	Rui Pedro dos Santos PatrÃ-cio		2.0	39.0	2.0	40.0	None
Iran	Iran	Mehdi Taremi		112.0	33.0			None
Iran	Iran	Mehdi Taremi	Left Foot	113.0	33.0	120.0	35.5	Off T
Iran	Portugal	Rui Pedro dos Santos PatrÃ-cio		4.0	45.0	5.0	45.0	None

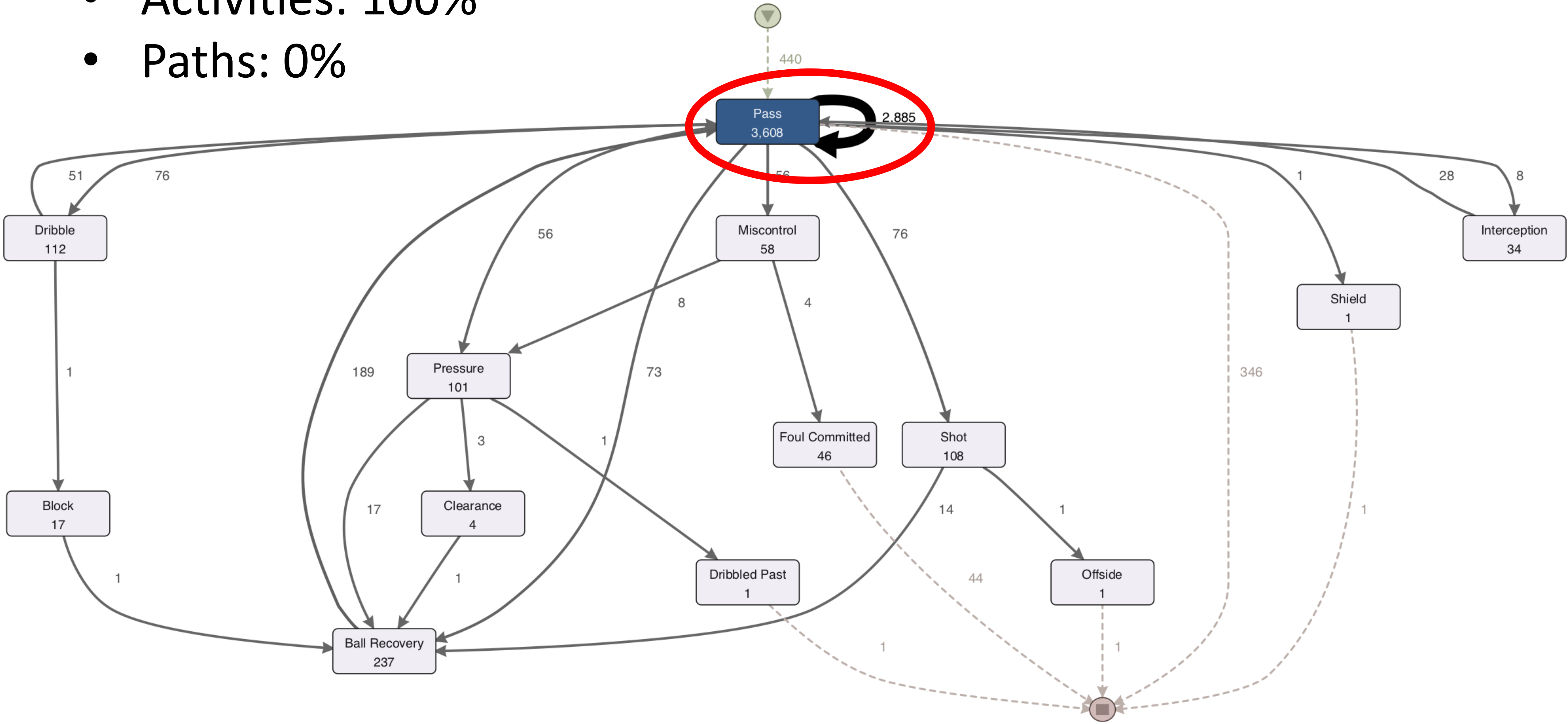
Belgium

7 matches

https://cdn.dnaindia.com/sites/default/files/styles/full/public/2018/07/07/701969-belgium-afp-national-team-2.jpg

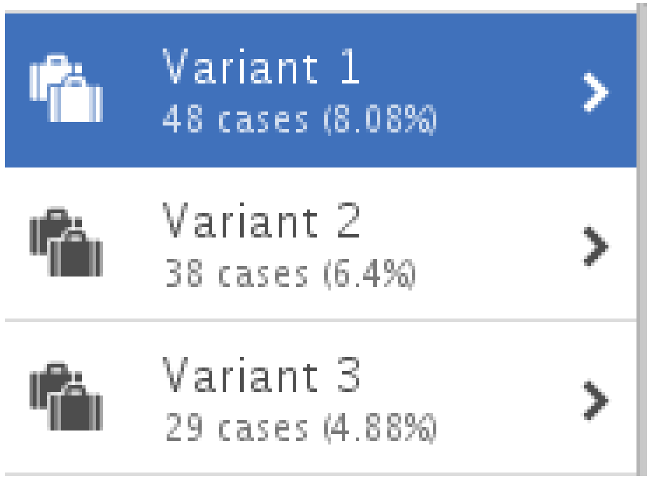
Disco

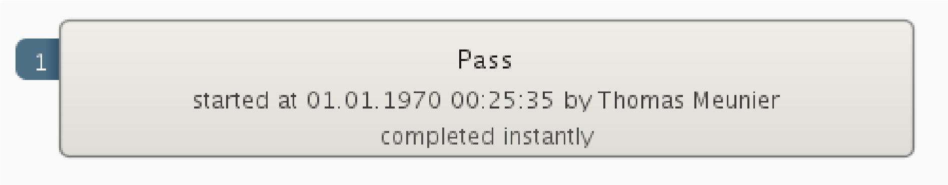
- Activities: 100%
- Paths: 0%

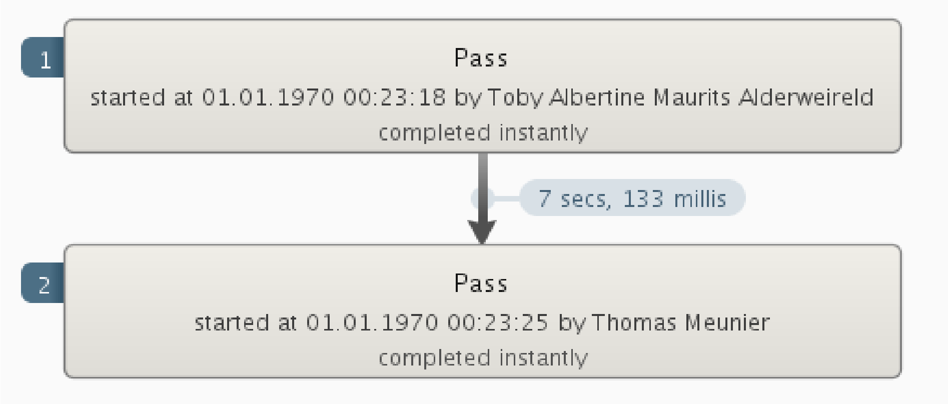


Statistics

- 594 cases
- 295 variants

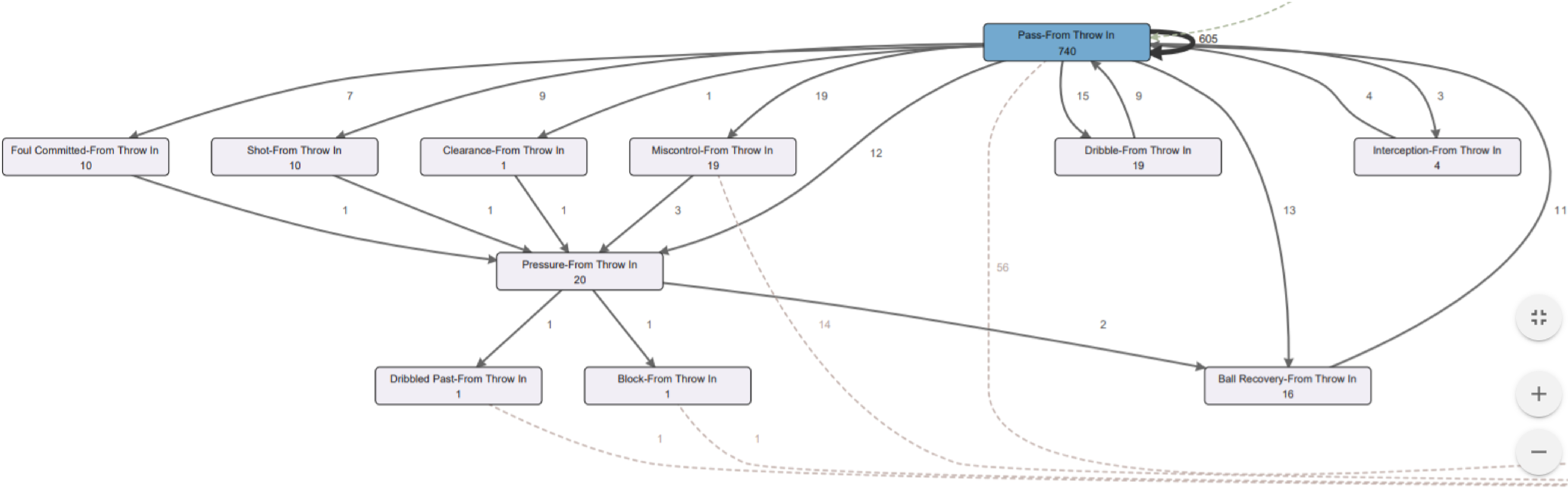




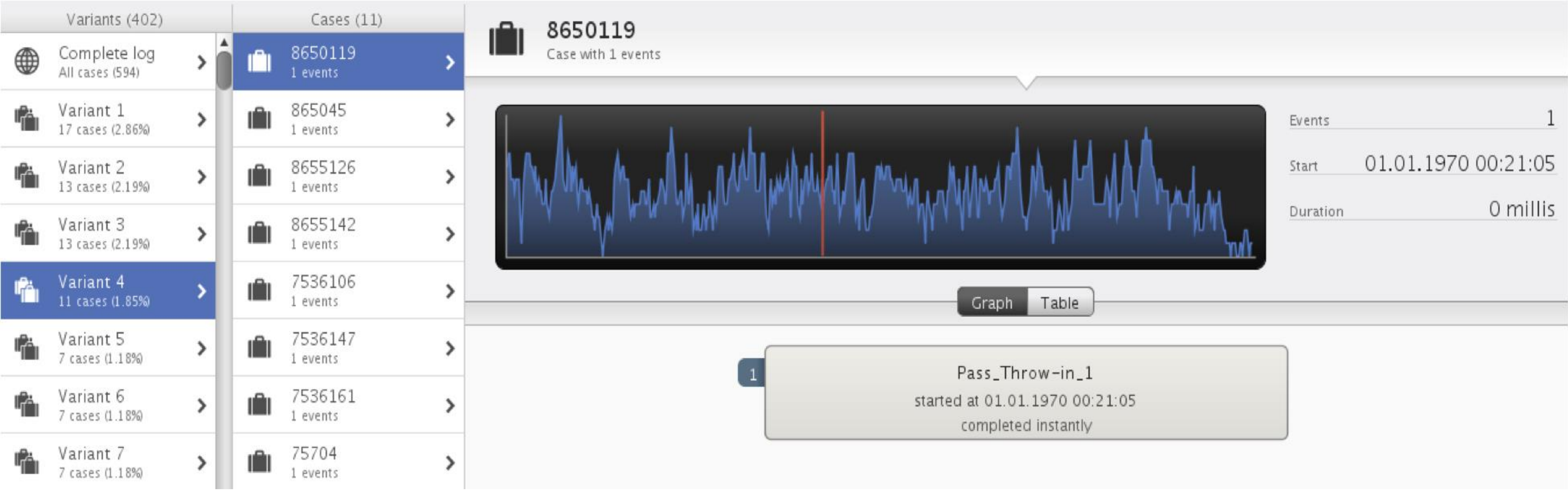


Set Pieces (sub-processes)

- Action + Type



Undesired cases



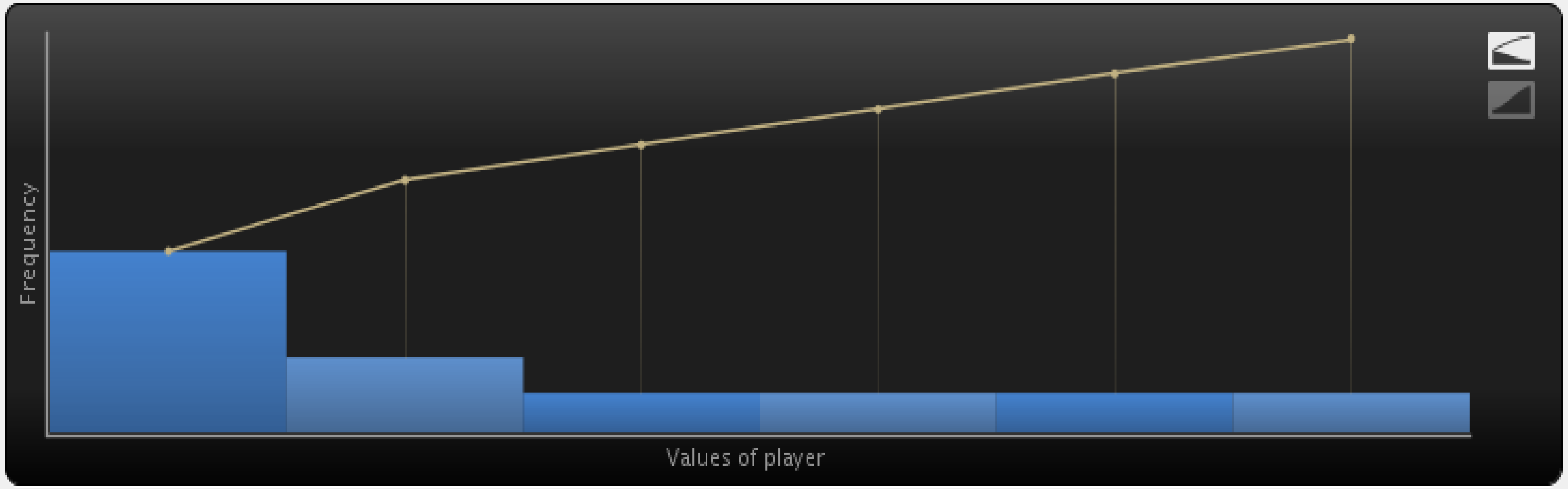
Undesired cases

Statistics views

- period — Other attribute
- minute — Other attribute
- second — Other attribute
- duration — Other attribute
- endTime — Other attribute
- possession_team — Other attribute
- team_action — Other attribute
- player — Other attribute
- player_position — Other attribute
- body_part — Other attribute

player

Other attribute

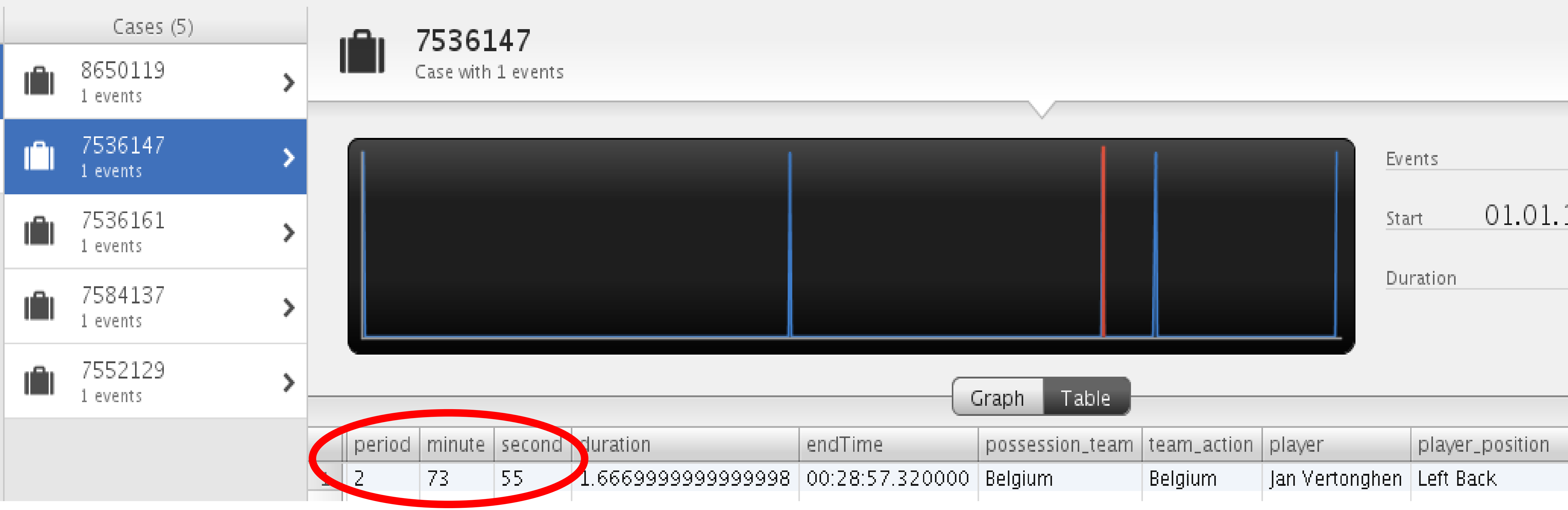


Values	6
Minimal frequency	1
Mean frequency	1.83
Maximal frequency	5
Frequency std. deviation	1.6

All values (6) | First in case (6) | Last in case (6)

Value	▲ Frequency	Relative frequency
Jan Vertonghen	5	45.45 %
Yannick Ferreira Carrasco	2	18.18 %
Thomas Meunier	1	9.09 %
Toby Albertine Maurits Alderweireld	1	9.09 %
Dries Mertens	1	9.09 %
Thomas Vermaelen	1	9.09 %

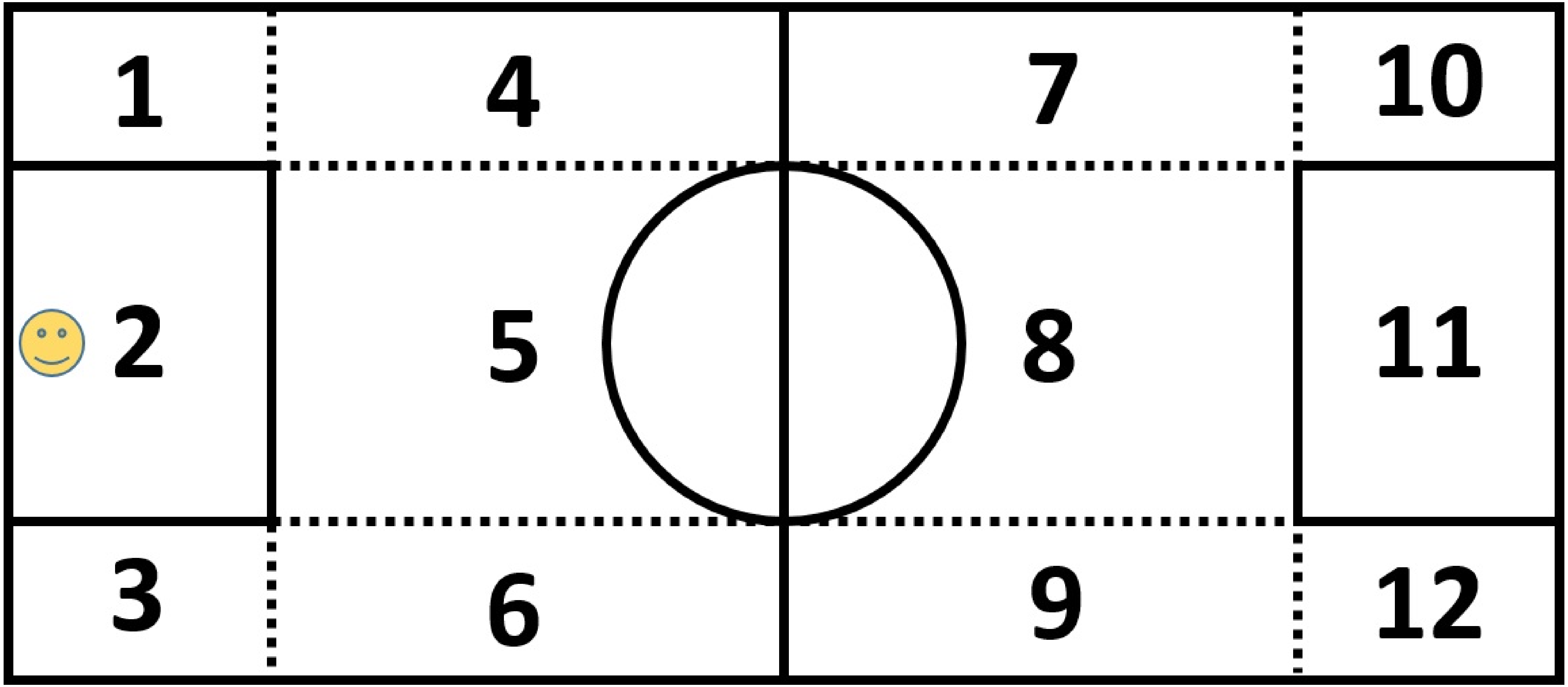
Undesired cases



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

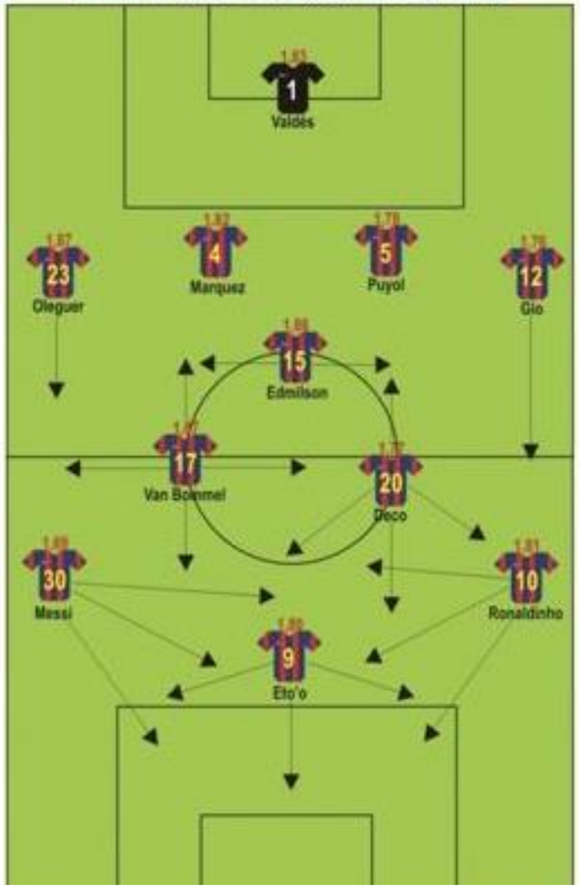

SPORTS
ANALYTICS
CHALLENGE
PARIS
SAINT-GERMAIN
ÉCOLE
POLYTECHNIQUE
Fly

15 years ago

https://secure.i.telegraph.co.uk/multimedia/archive/02186/mourinho_lampard_2186242b.jpg

15 years ago

Probable system of play and starting line-up



Offensive Organization

* Team organized in a pure 4x3x3. From last year have improved in the agressiveness of their offensive transition but on the other hand are much weaker in that characteristical 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 possi_ ble. 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 are more pl_ ayers 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 provocates between the lines it's important to communicate with deffensive midfielder because the positioning can be too 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 stoped 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 o_ pposite 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 Betis) coming inside on an agressive 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 antecipates and reads the off-side line- depth.
* Eto'o alternates his positioning in relation to the positioning of the ball. In the build-up he will come short between lines tobring defender out, then touches and goes in behind. In the final phase he puts himself between defenders positioning and away from marker, then diagonals!!!

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

https://twitter.com/AnalysisLab1/status/1106722625470889984

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

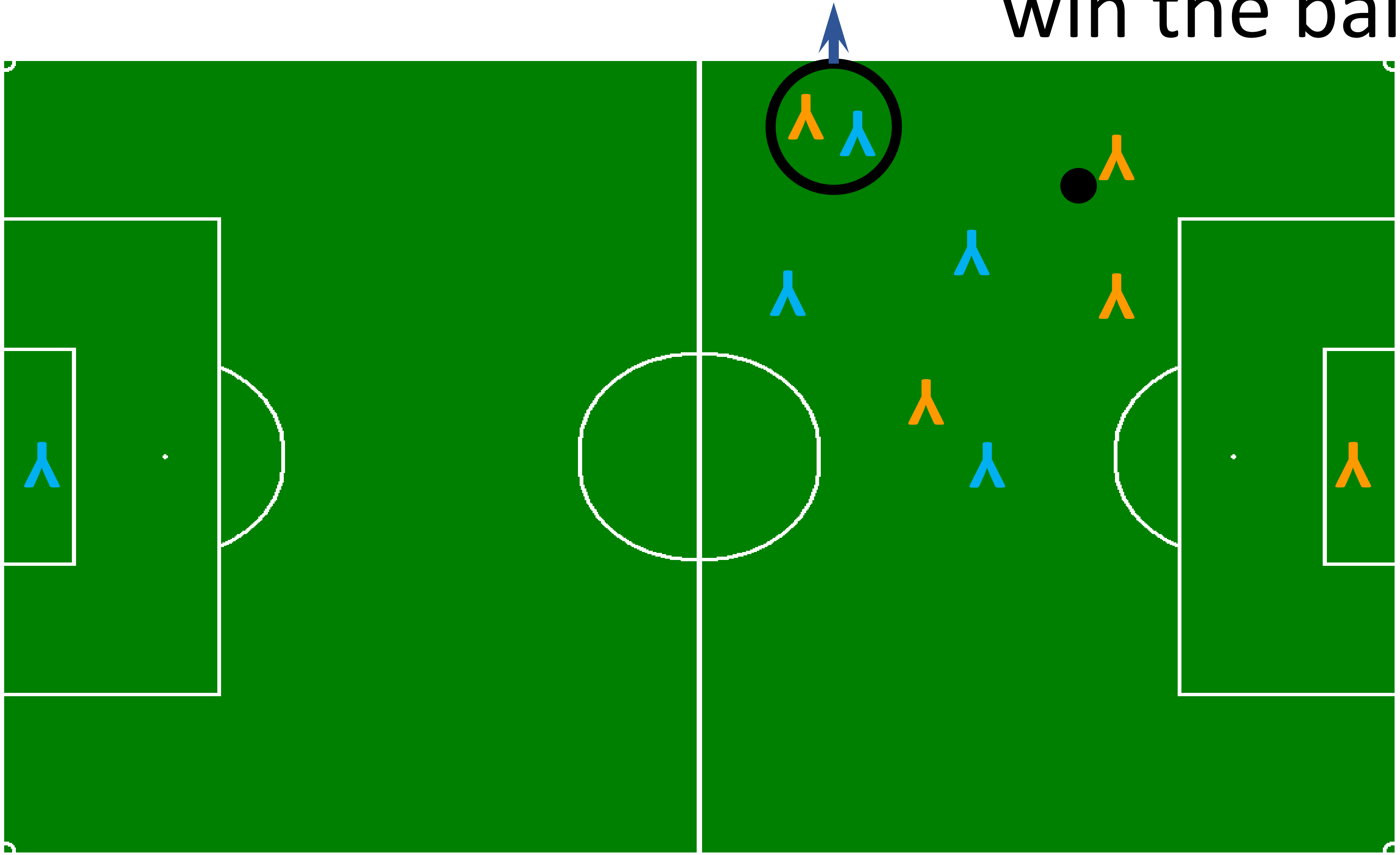
Pressing

Success

≤3 meters

win the ball within 5s

Team A



Team B

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

Approach

Machine Learning:

$\mathbf{X}$ = Pressing start moment

$\mathbf{Y}$ = Successful or Unsuccessful

$$1)\ f(\mathbf{X}) = \mathbf{Y}$$

$$2)\ g(\mathbf{X}) \simeq f(\mathbf{X})$$

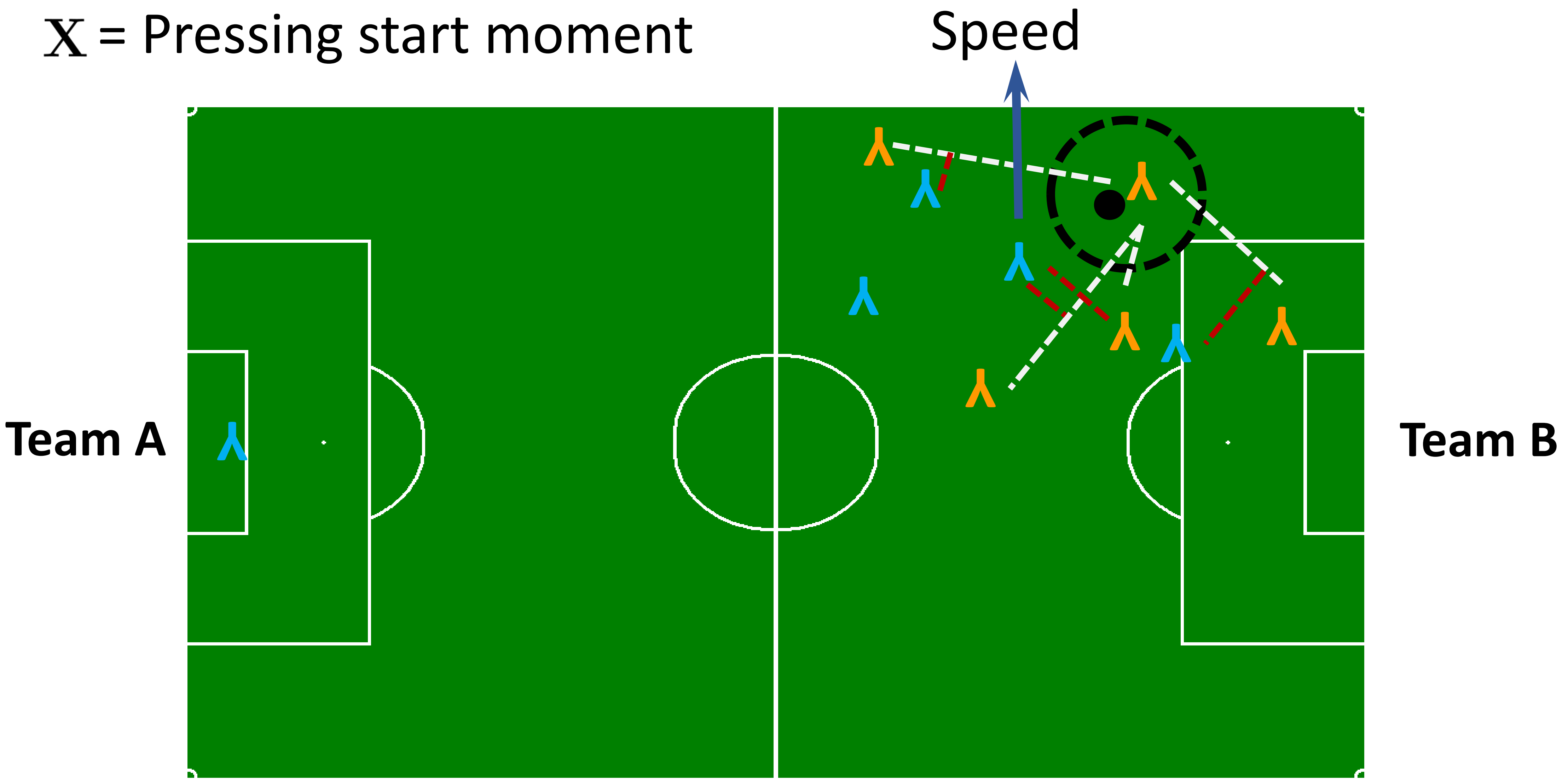
Pattern Extraction

1) $$g(\underbrace{\mathbf{x}_1}_{\text{Increase/decrease}}, \overbrace{\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_N}^{\text{Fixed}}) = \text{Prediction}$$

2) $$g(\underbrace{\mathbf{x}_1, \mathbf{x}_2}_{\text{Increase/decrease}}, \overbrace{\mathbf{x}_3, \ldots, \mathbf{x}_N}^{\text{Fixed}}) = \text{Prediction}$$

Attributes

X = Pressing start moment



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

Accuracy

Per Team:

1. Train/Test
- First 80% matches
- Last 20% matches

2. Algorithms
- Logistic Regression
- Decision Tree
- **Random Forest**
- Naïve Bayes
- Light Gradient Boosting

Lyon:

1. Accuracy
- Training: **81%**
- Test: **71%**

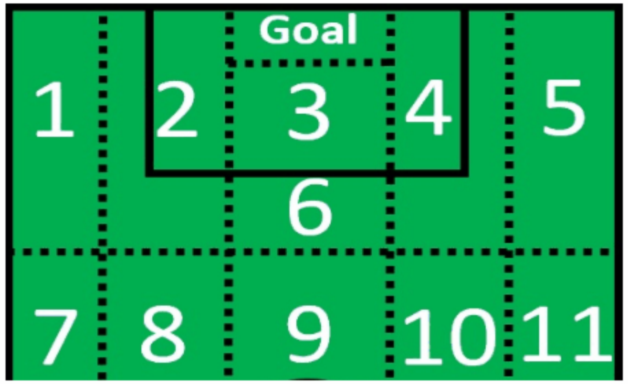
2. Baseline
- Majority Class : **55%**

How to prepare for a new match?

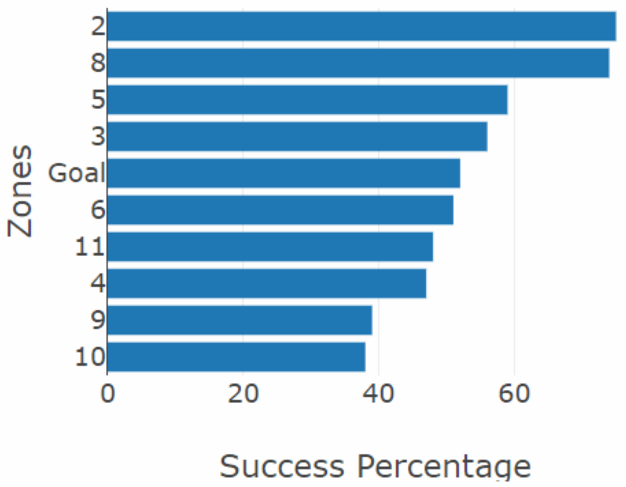
https://www.oddsfan.com/board/match-preview-psg-vs-lyon/

Overview | **Insight Engine** | **Case Explorer**

Zones



Target zones to put Lyon under pressure



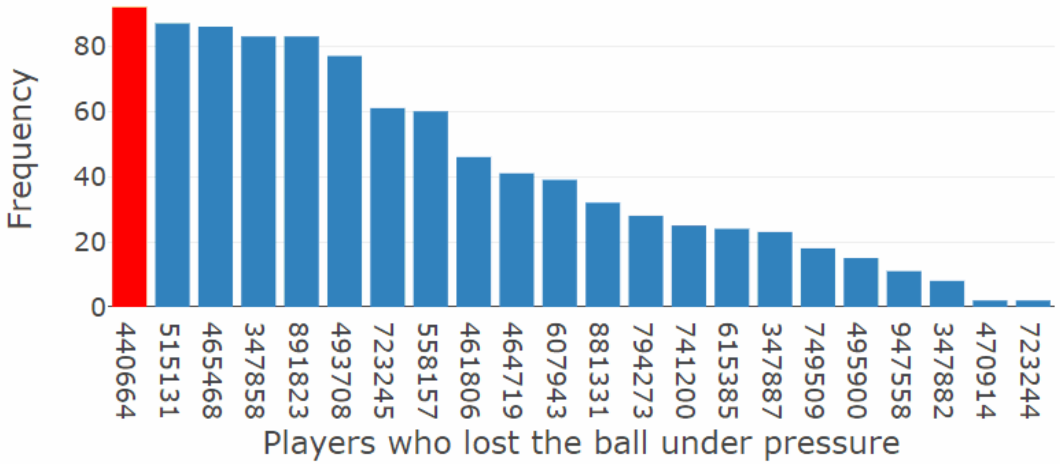
Team

Lyon

Number of Matches: 17

Pressure Success Rate Against Lyon : 57.78%

Players to put under pressure




PARIS SAINT-GERMAIN
RÊVONS PLUS GRAND


ICI C'EST PARIS
RÊVONS PLUS GRAND


SOTUDEH
2
ooredoo

Thank you for your attention

h.sotudeh@tue.nl

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