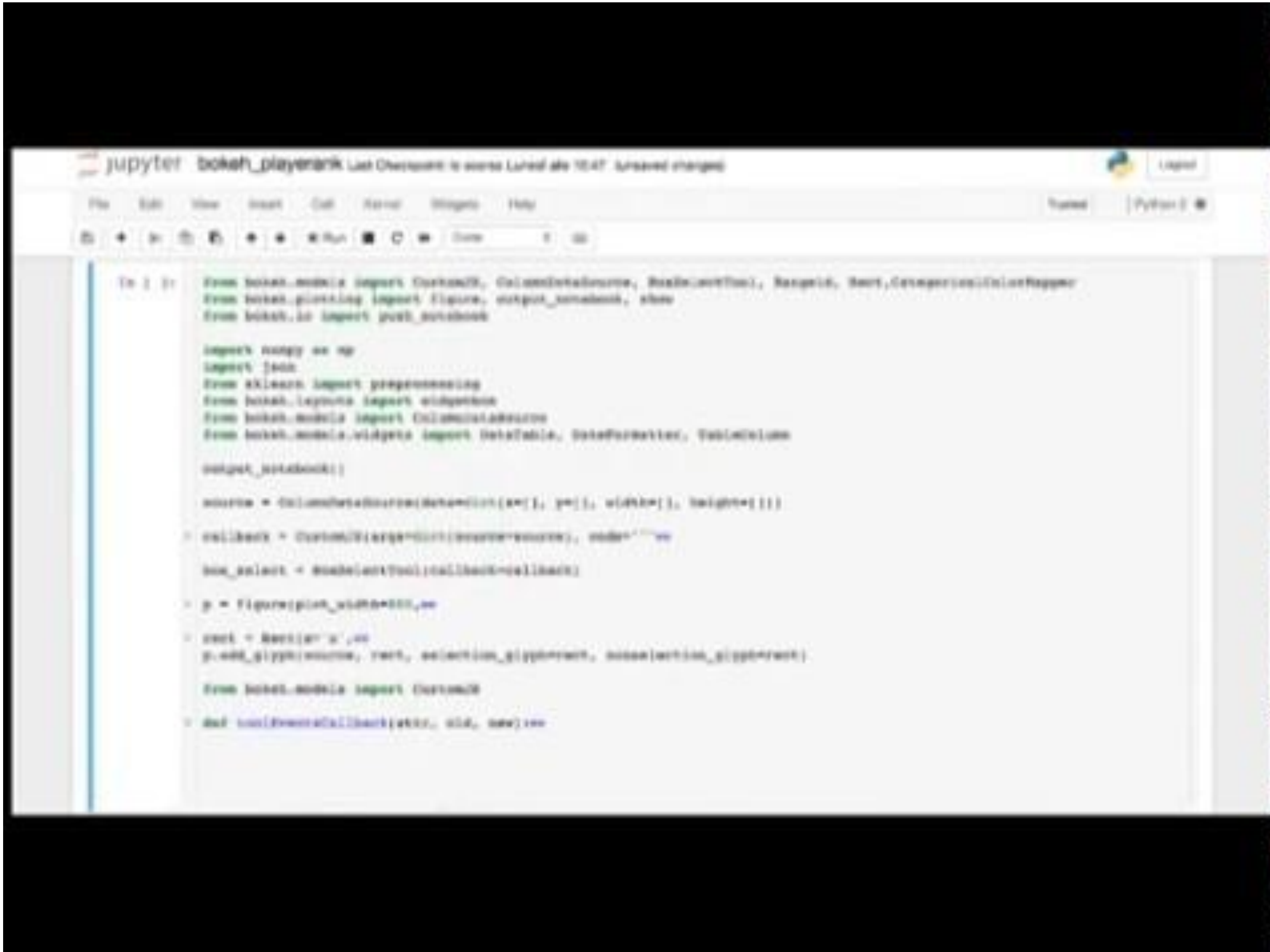
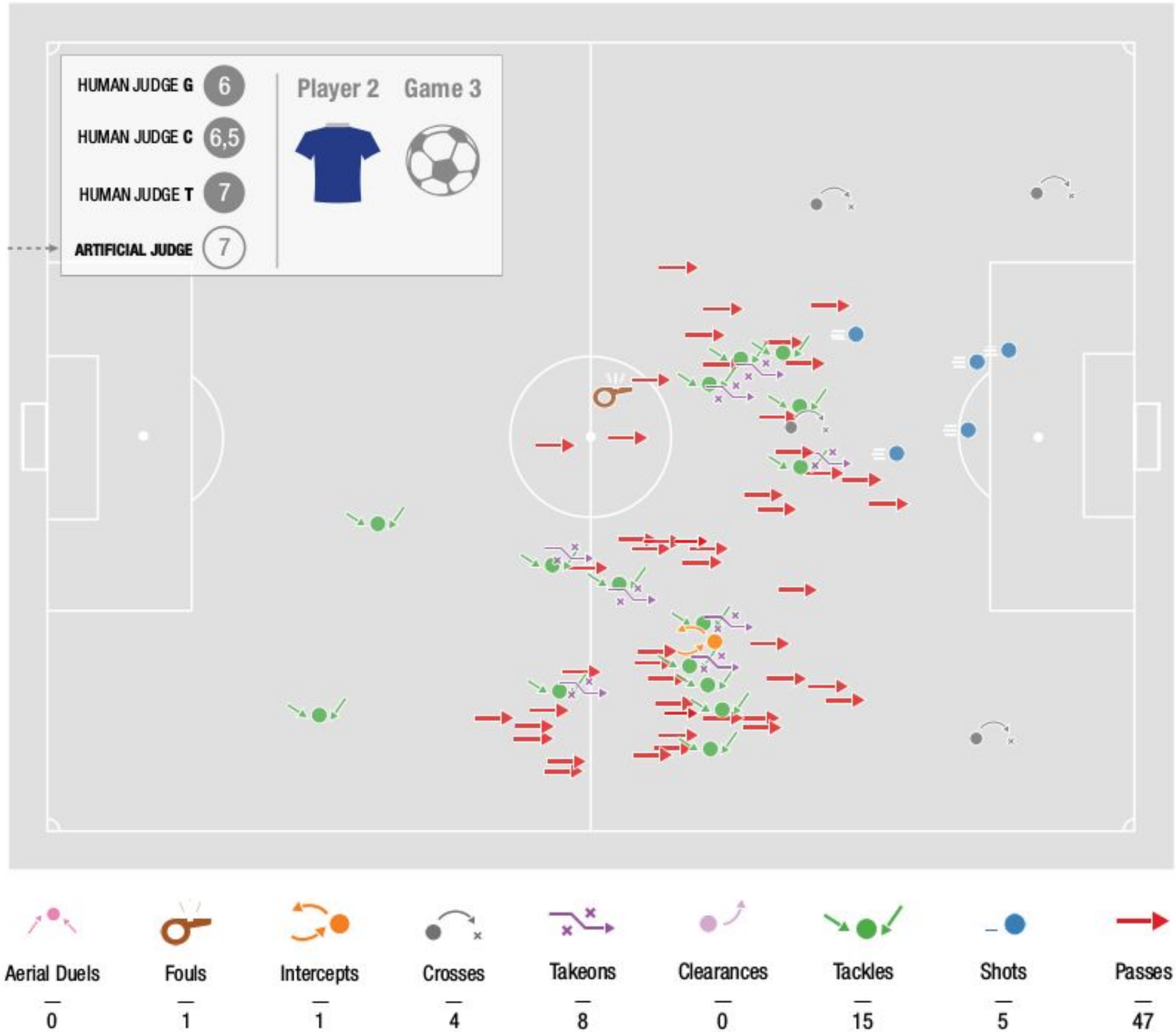
Performance evaluation in soccer

from **human** mechanisms to **data-driven** algorithms


jupyter
bokeh_playerank


HUMAN JUDGE G 6
HUMAN JUDGE C 6,5
HUMAN JUDGE T 7
ARTIFICIAL JUDGE 7
Player 2
Game 3
Aerial Duels 0
Fouls 1
Intercepts 1
Crosses 4
Takeons 8
Clearances 0
Tackles 15
Shots 5
Passes 47

```
{'eventName': 8,
  'eventSec': 8.221464,
  'id': 217097515,
  'matchId': 2576132,
  'matchPeriod': '1H',
  'playerId': 8306,
  'positions': [{'x': 42, 'y': 14},
{'x': 74, 'y': 33}],
  'subEventName': 83,
  'tags': [{'id': 1801}],
  'teamId': 3158}
```

pass

identifiers

high pass

accurate

mongoDB®

1700 events per match (in average)

powered by wyscout

pass

1700 events per match (on average)


```
{'eventName': 8,
  'eventSec': 8.221464,
  'id': 217097515,
  'matchId': 2576132,
  'matchPeriod': '1H',
  'playerId': 8306,
  'positions': [{'x': 42, 'y': 14}, {'x': 74,
'y': 33}],
  'subEventName': 83,
  'tags': [{'id': 1801}],
  'teamId': 3158}
```


identifiers

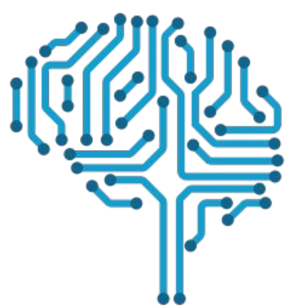
accurate

Performance vector

passes	xG	pressing	accuracy	...	...	...	...	...	...	...	...

How to *automatically* evaluate performance?

solution:
imitate the human



Soccer Player Ratings

LE PAGELLE di Antonio Giordano

ZIELINSKI, EREDE DI QUALITÀ


6

REINA
Per i fantacalcisti e perché sull'unico pallone rischia la salute andando nella pozzanghera.


6

HYSAJ
Eppure il campo non gli manca (non gli mancherebbe) ma le energie forse un pochino sì.


6

ALBIOL
Con le ciabatte, in stile salotto, lasciando che la Spal gli vada a battere addosso.


6,5

KOULIBALY
Il solito «energumeno»: di forza, di prepotenza e con autorevolezza ritrovata.


6

MARIO RUI
Rischia il giallo (e la squalifica) e quindi poi si contiene, limitandosi.


6,5

MERET
E' bravo, reattivo, istintivo e frena Insigne ma soprattutto Callejon.


5,5

SALOMON
Non sceglie: aspetta o attacca Insigne e rischia di finire a gambe all'aria.


6

VICARI
Sta là dietro e oppone il corpo e la posizione alle rare verticalizzazioni.


5,5

FELIPE
Si stacca troppo, aprendo la corsia centrale per Allan, perché Callejon lo distrae.


5,5

LAZZARI
Gli mancano le coperture e poi dà un senso di anarchia tagliando sempre, troppo.


6,5

ALLAN
Il gol che riconsegna il primato in classifica, prima di correre per sé e per gli altri.


6

JORGINHO
Geometrie apprezzabili, però senza avere intorno uomini che pedalino come si dovrebbe.


5,5

HAMSIK
Il pallido capitano rimane dietro i suoi standard e l'ammonizione gli fa male.


6

CALLEJON
Apre per Allan e lo manda in porta e poi (sembra) governa i carichi di fatica.


6

MERTENS
E' la prima sponda nell'1-0 ma è anche un po' vago, quasi distante dalla partita.


6

SCHIATTARELLA
Si ritrova con Hamsik, lo contiene e persino lo costringe a stargli dietro.


5,5

VIVIANI
Gli viene meno il gusto di osare e palleggia con paura addosso che diventa nemica.


6

GRASSI
Perde lo scatto di Allan, poi dà movimento e pure eleganza ad un centrocampo piatto.


5

DRAMÈ
Quasi si isola e lascia che da quelle parti, ma senza esagerare, il Napoli vada.


5

KURTIC
L'unica preoccupazione è Jorginho e spreca non l'occasione ma il suo tempo.


6

INSIGNE
Insegue il gol, e si vede, però Meret e il palo lo costringono a soffrire ancora.


6,5

ZIELINSKI
(25' st)
E' di impatto ma anche di talento (e che ruleta!). Hamsik ha un erede di qualità assoluta.


sv

ROG
(41' st)
Va a coprire il campo, per restringerlo, nel finale da domare con intelligenza.


sv

DIAWARA
(45' st)
L'ultimo argine per il recupero che diventa ampio e comunque pericoloso.

6

SARRI
Piccole tracce di Napoli, qualcosina all'avvio, poi una gestione eccesiva.

5,5

SEMPLICI
Magari un pizzico di coraggio in più, solo quello, per dire di averci provato.


5,5

ANTENUCCI
Non gli arriva uno straccio di pallone, ma non ne va neanche a inseguire.


5,5

COSTA
(16' st)
In un contesto blando a cui può solo garantire di fungere da cerniera.


sv

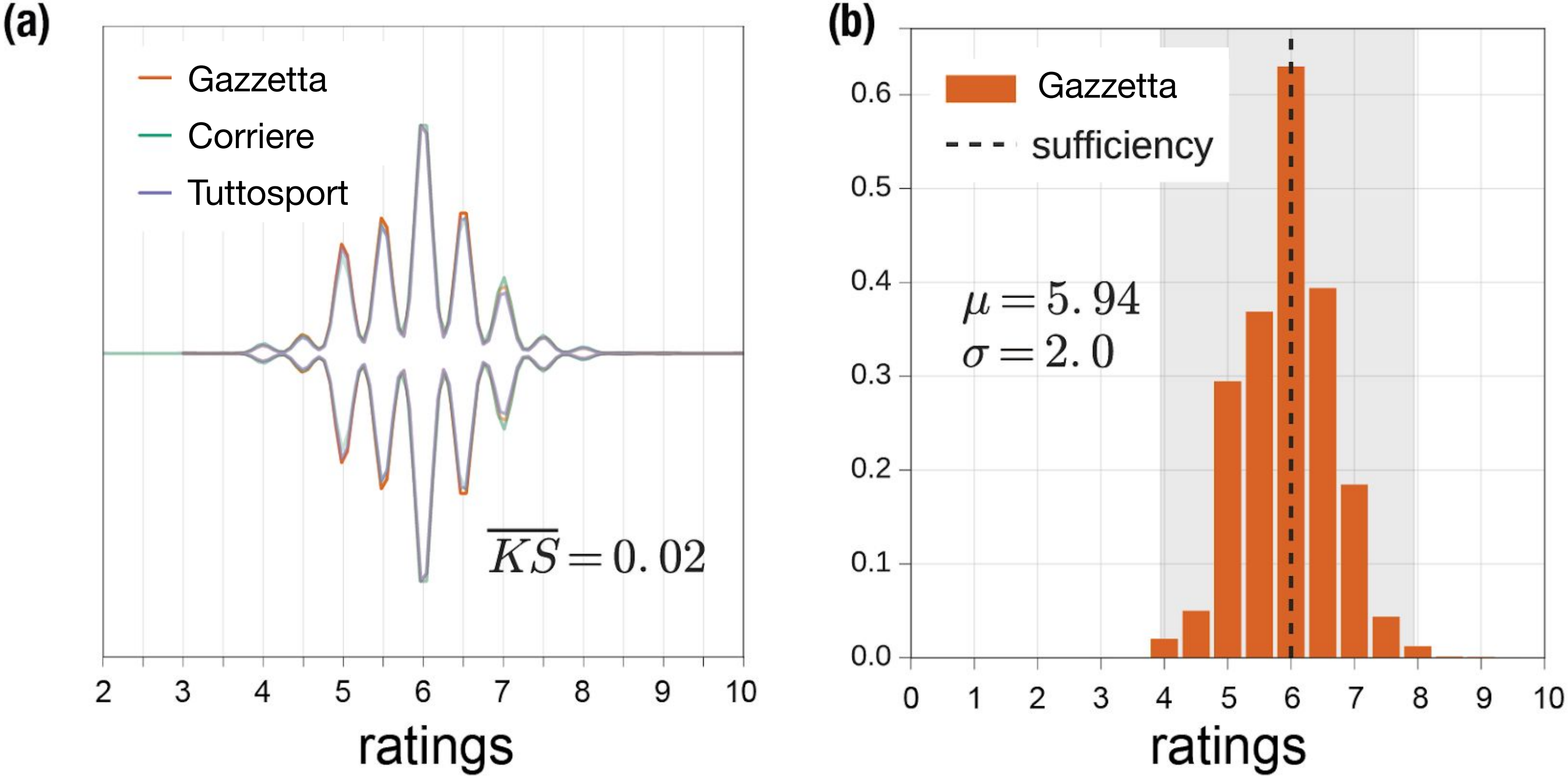
FLOCCARI
(30' st)
E' il jolly che si va a cercare: magari una palla sporca. Ma bisognerebbe arrivare a lui.


sv

PALOSCHI
(37' st)
Aggiunge spiccioli di minutaggio ad una gara in cui l'attacco non esiste.

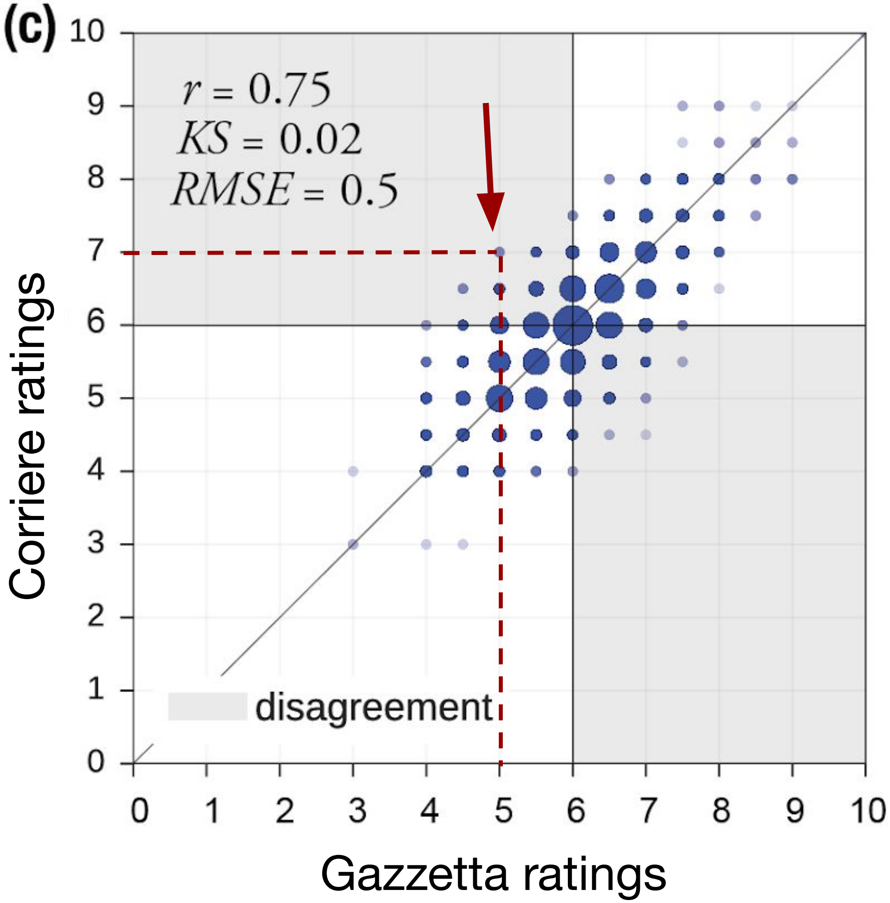

6

GAVILLUCCI
Già non averla complicata, semplice com'era, sa di buon senso. Comoda così eh



- identical distributions
- peak at sufficiency
- extreme ratings are rare


Gonzalo Higuain, 28 anni, tre gol ieri sera contro il Frosinone: sono 36 totali
36
IL PIPITA RE DEL GOL
BATTUTO LO STORICO
RECORD DI NORDAHL
IL NAPOLI È SECONDO
SUA MAESTÀ
HIGUAIN
Tris al Frosinone, superato lo svedese che
nel '50 segnò 35 reti. Roma bella ma terza
Il Milan umiliato. Sassuolo sesto, Inter ko
SERIE A 38ª GIORNATA
di Antonio Giordano
(b)
G
sufficiency
$\mu = 5.94$
$\sigma = 2.0$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0 1 2 3 4 5 6 7 8 9 10
ratings
10



Statistical agreement

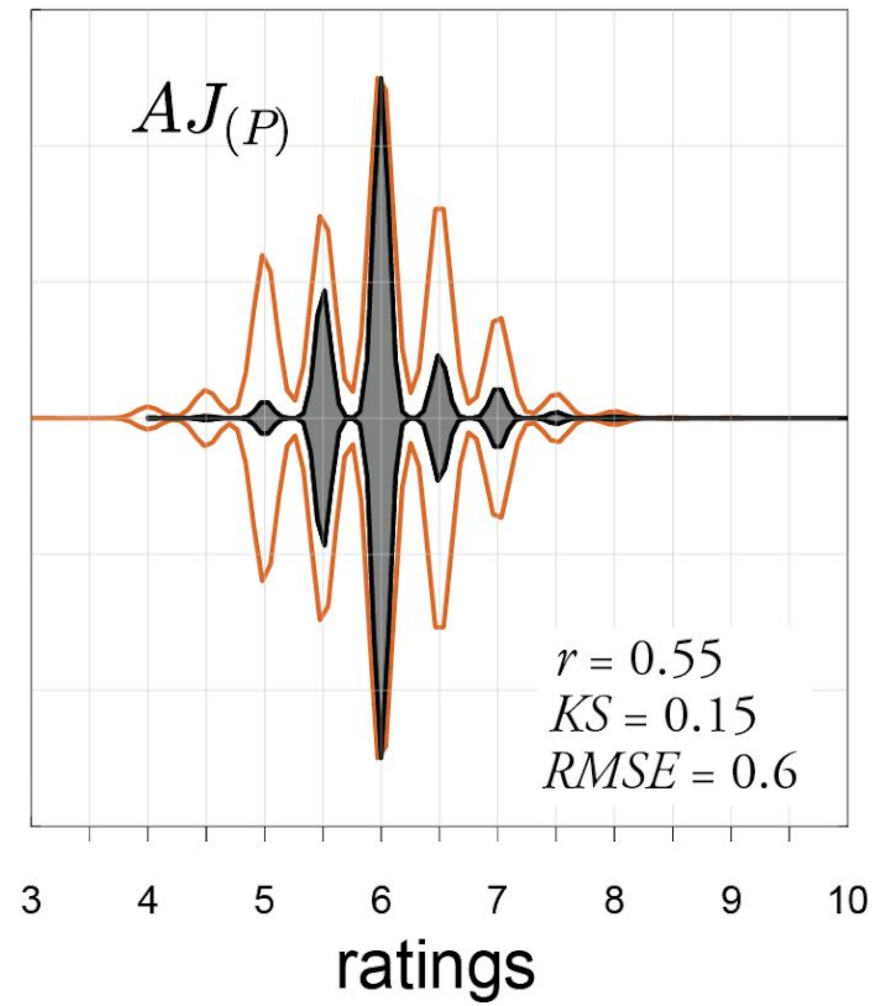
correlation of two judges $r=0.75$

- judges often agree...
- ...but areas of disagreement exist

We use performance and ratings to create an

artificial judge $AJ_{(P)}$

to predict ratings from performance

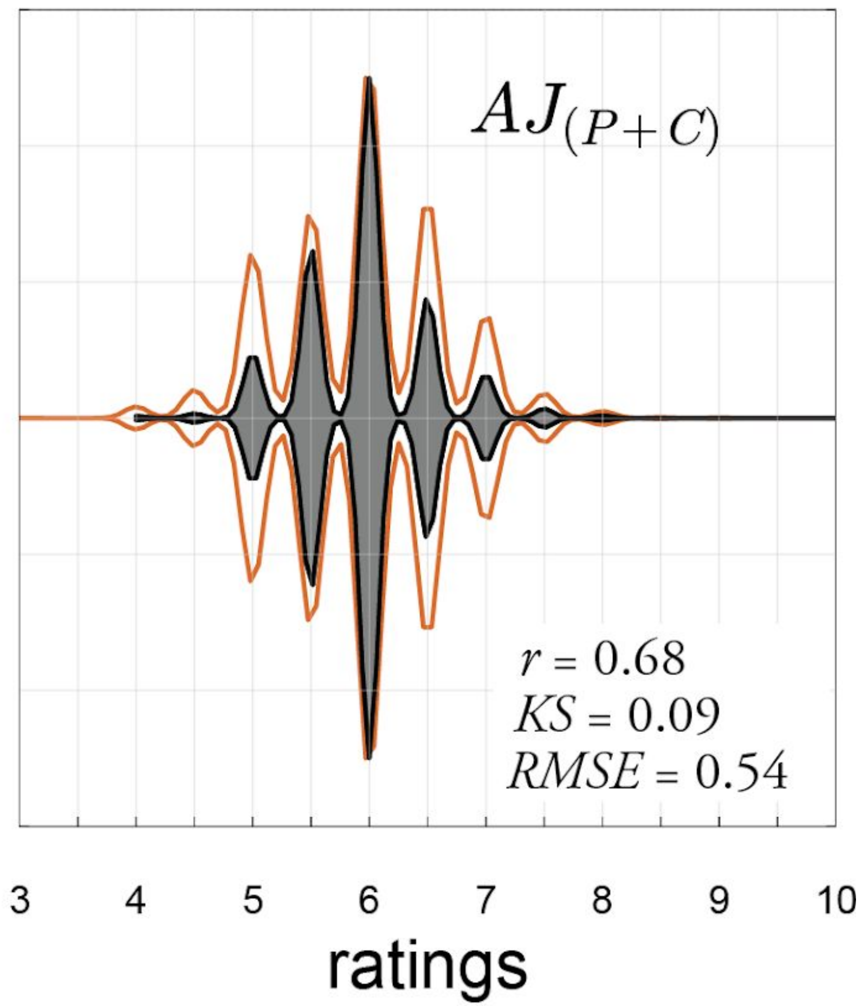


r = 0.55

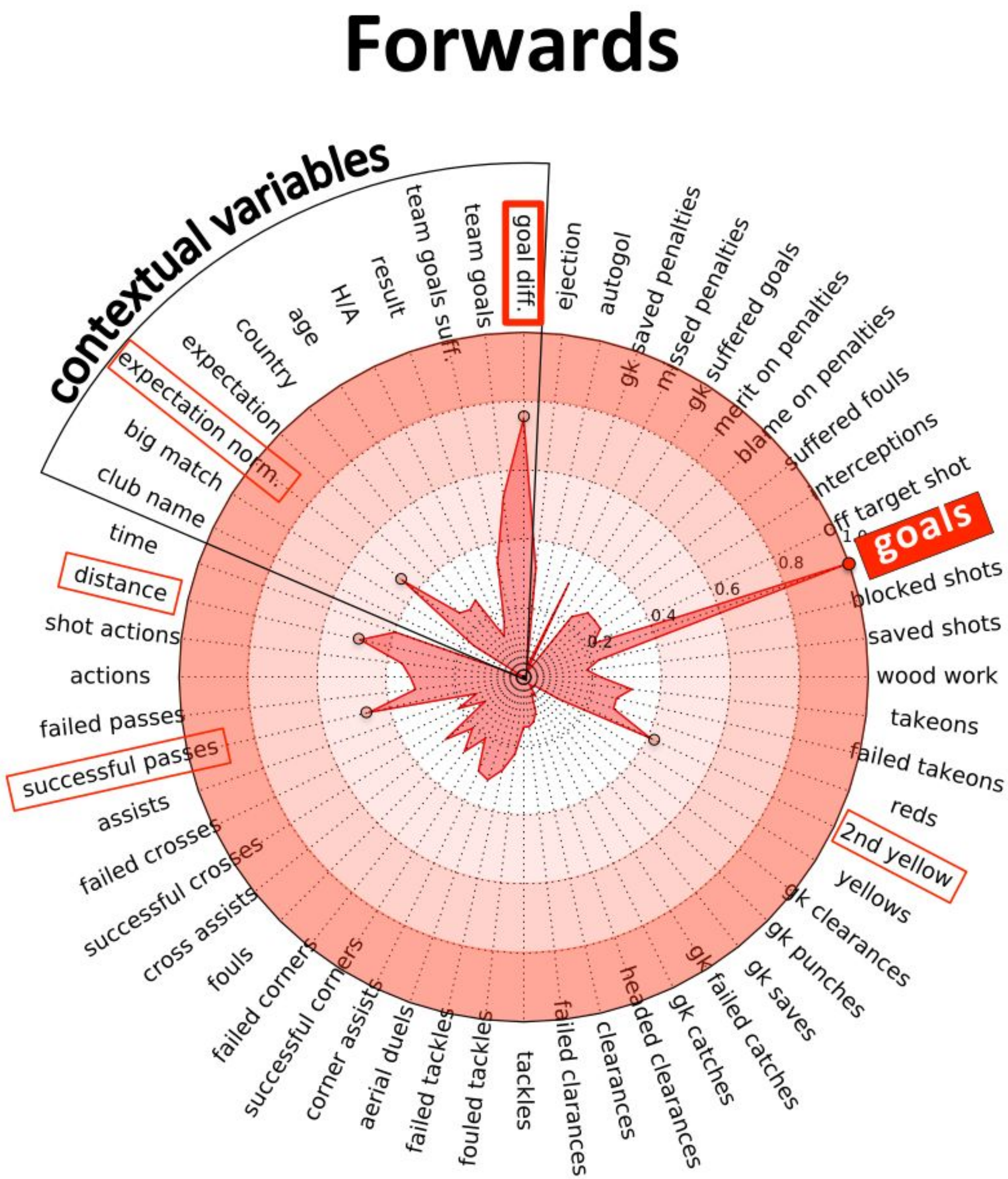
Home /Away	Quote	Game outcome	Age	Big match

We add ***contextual*** info to create an

artificial judge $AJ_{(P+C)}$



$\boldsymbol{r} = 0.55 \rightarrow \mathbf{0.68}$



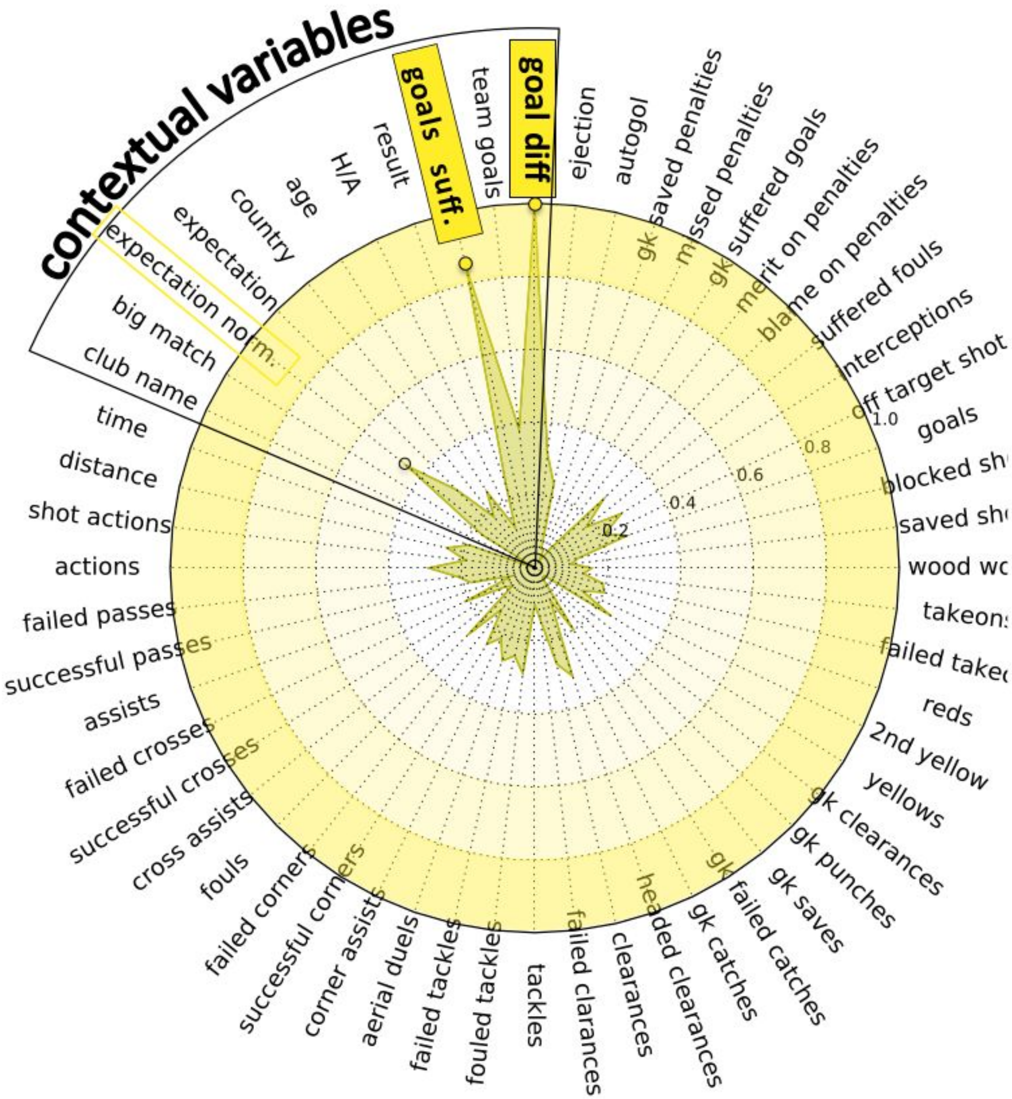
Features that matter

1) Just a subset of the features matter (20)

2) Contextual features are highly important

3) >90% of the features have negligible importance

Defenders



Features that matter

1) Just a subset of the features matter (20)

2) Contextual features are highly important

3) >90% of the features have negligible importance

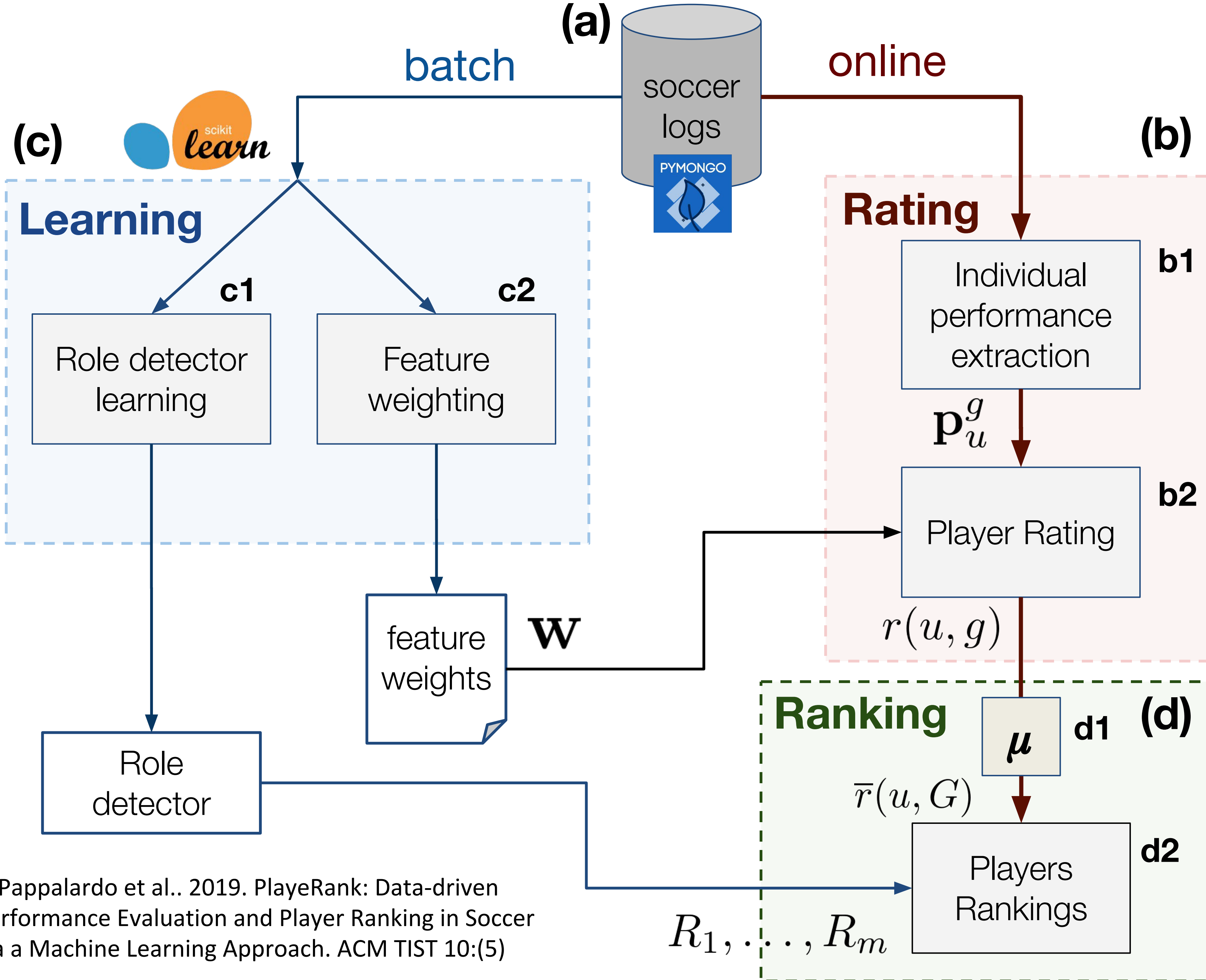
4) the same features has different importance in different roles

How to *automatically* evaluate performance?

solution:

~~imitate the human~~

make it data-driven



L. Pappalardo et al.. 2019. PlayeRank: Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach. ACM TIST 10:(5)

Step #1: player performance

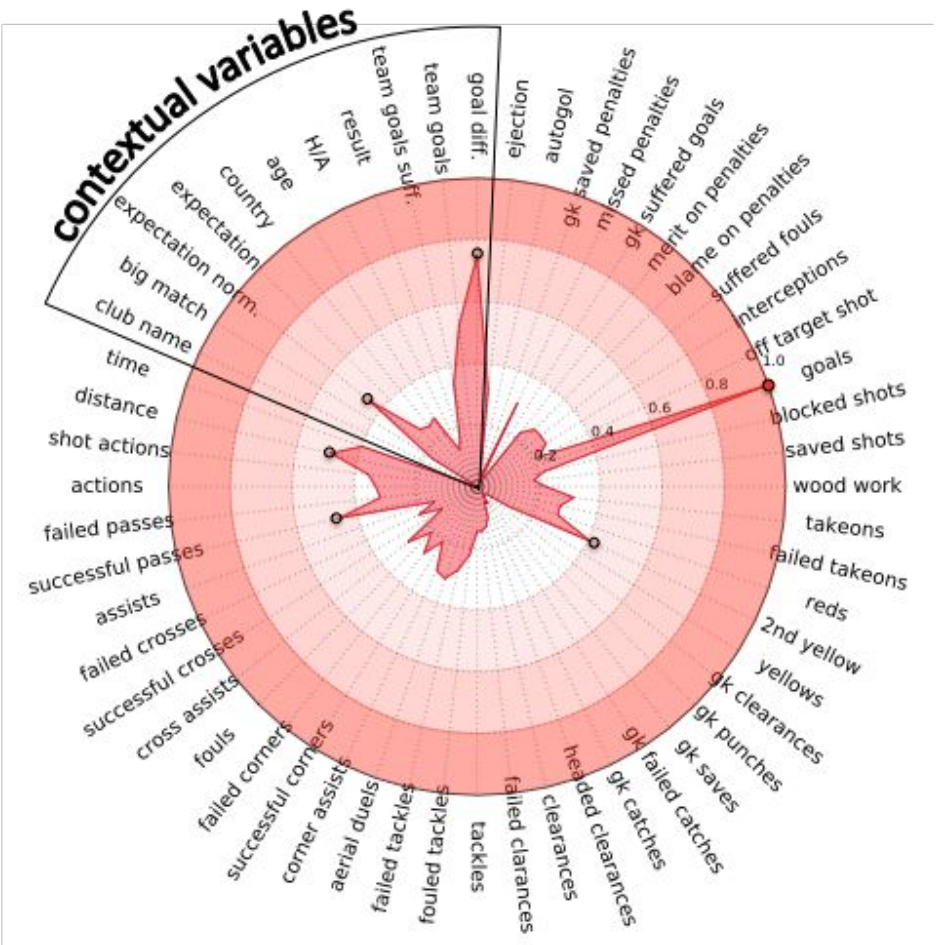
14 million events

150 technical features

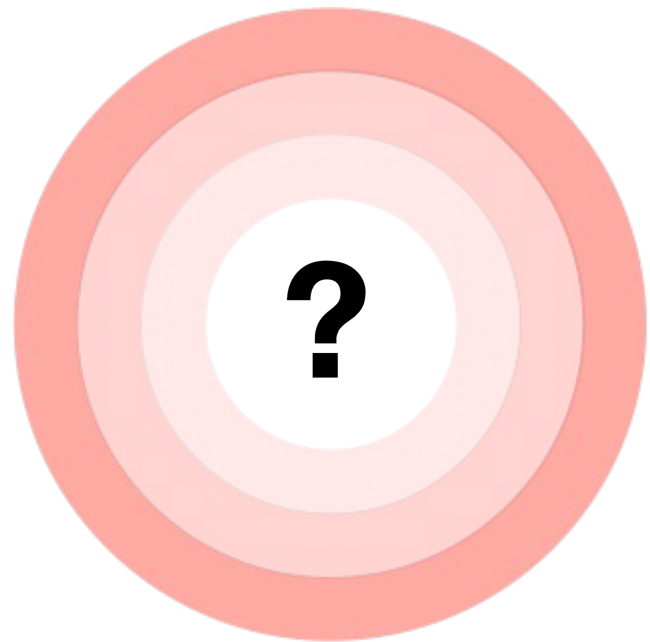
7,304 games

1,192 professional players

Step #2: feature weighting

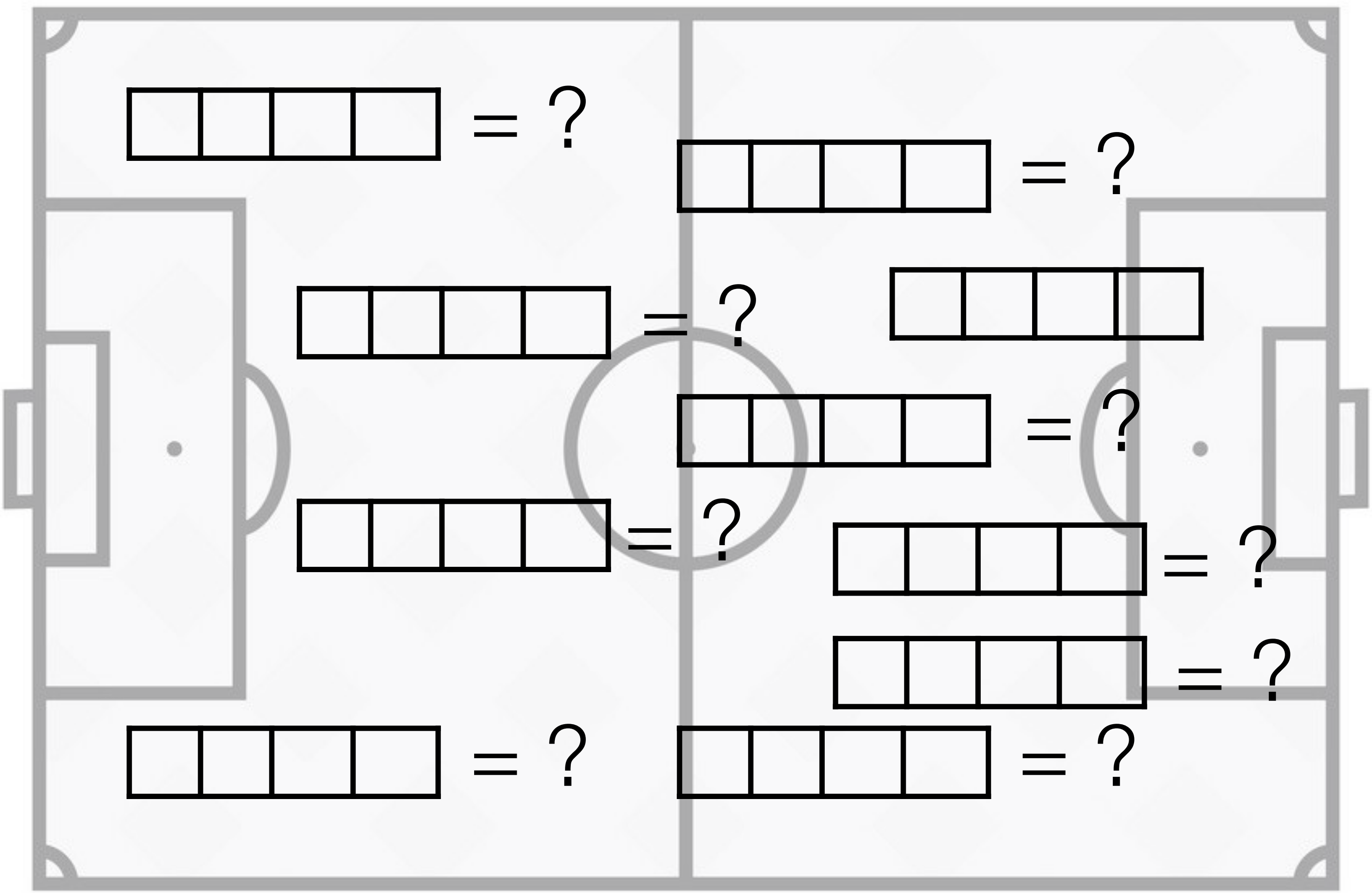


Human rating

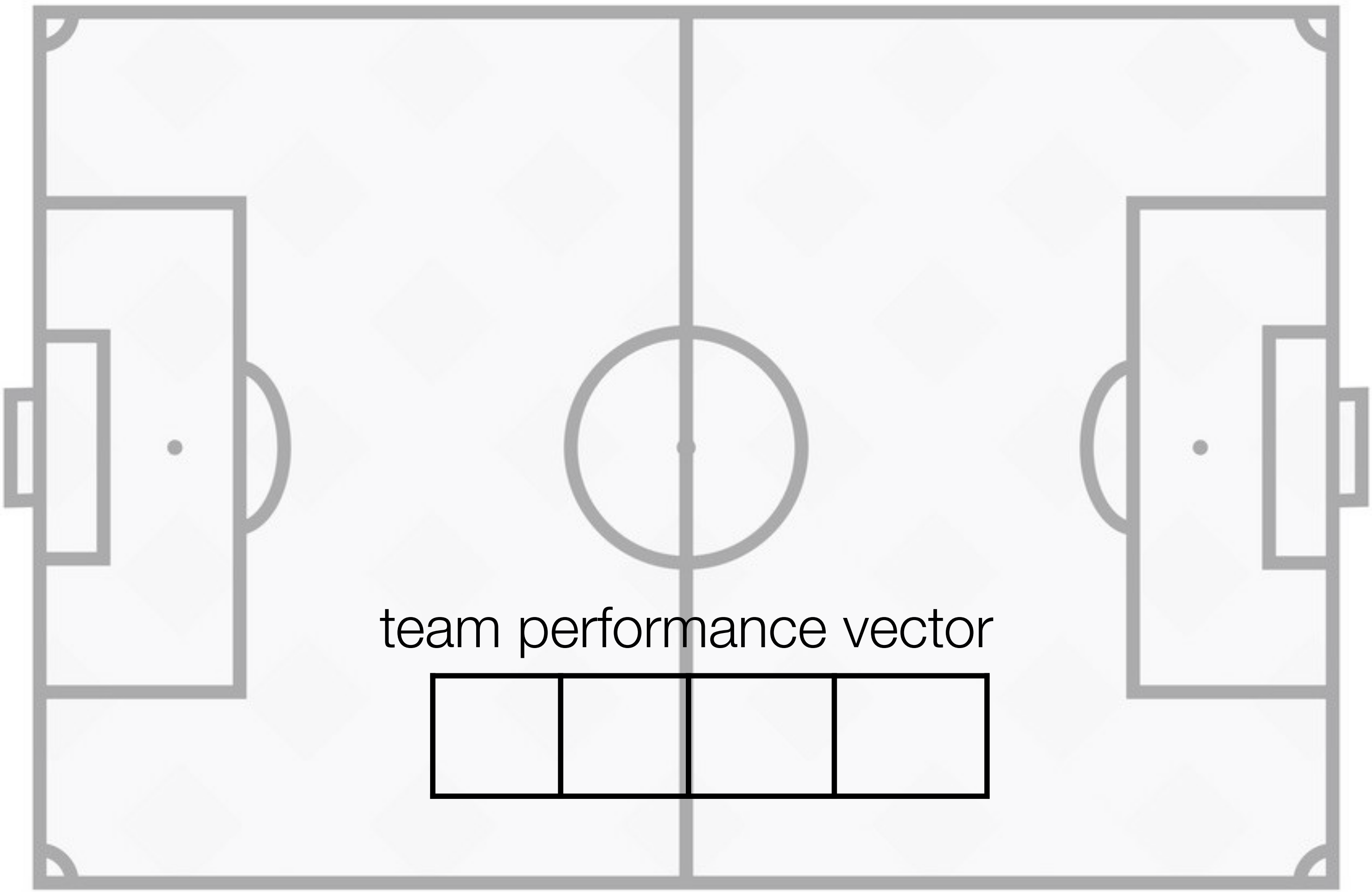


Data-driven rating

Feature Weighting



Feature Weighting



Feature Weighting

team1

passes	xG	pressing	accuracy	...

?

1X2

team2

passes	xG	pressing	accuracy	...

Pappalardo and Cintia, (2017) Quantifying the relation between performance and success in soccer, Advances in Complex Systems, doi:10.1142/S021952591750014X

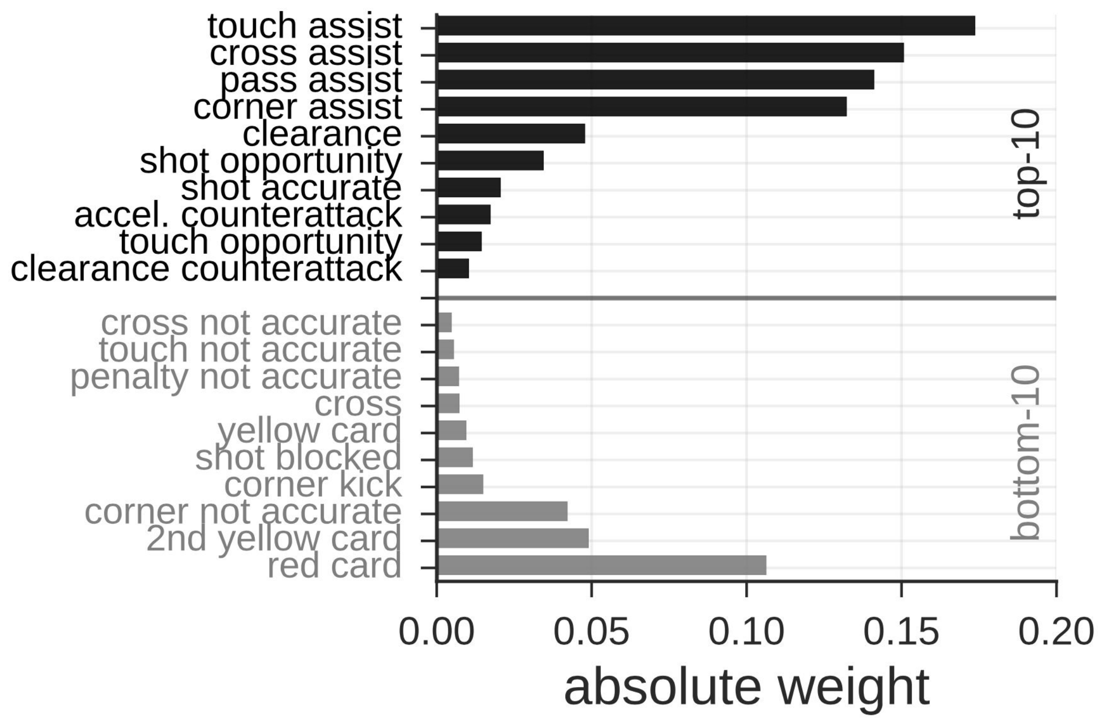
Step #2: feature weighting

```
from pymongo import MongoClient

client = MongoClient('localhost', 27017)
events = client.wyscout.events

res = events.map_reduce(map_agg, reduce_sum)
X, y = extract_data(res)
```

	inaccurate defending duel	intercept	accurate air duel	accelleration	corner assist	missed penalty	foul	corner pass	accurate defending duel	cross key pass	...	outcome
0	-8.0	5.0	2.0	-1.0	0.0	0.0	-3.0	1.0	12.0	2.0	...	w
1	8.0	-5.0	-2.0	1.0	0.0	0.0	3.0	-1.0	-12.0	-2.0	...	n
2	-7.0	-3.0	6.0	0.0	0.0	0.0	5.0	-1.0	-10.0	1.0	...	w
3	7.0	3.0	-6.0	0.0	0.0	0.0	-5.0	1.0	10.0	-1.0	...	w
4	-13.0	-5.0	6.0	1.0	0.0	0.0	-6.0	1.0	-13.0	-2.0	...	w

Step #2: feature weighting

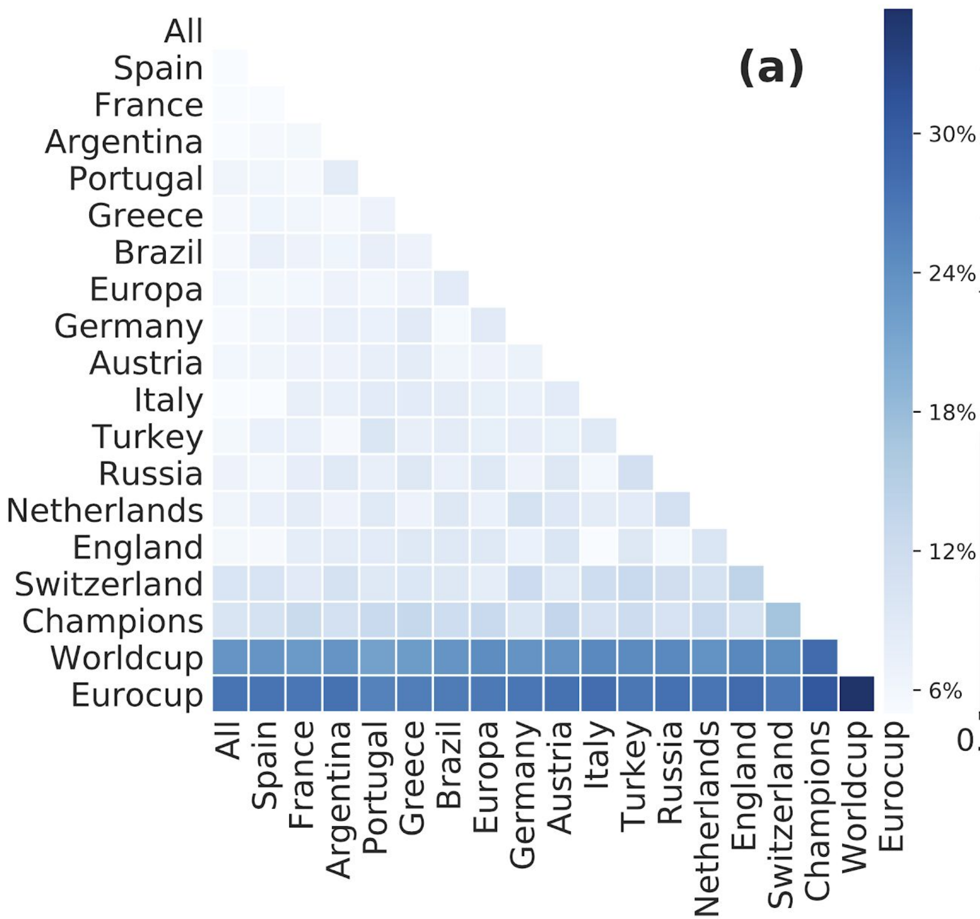
```
from playerank import Weighter
# perform the feature weighting
pw = Weighter()
pw.fit(X, y)
pw.weights_
```



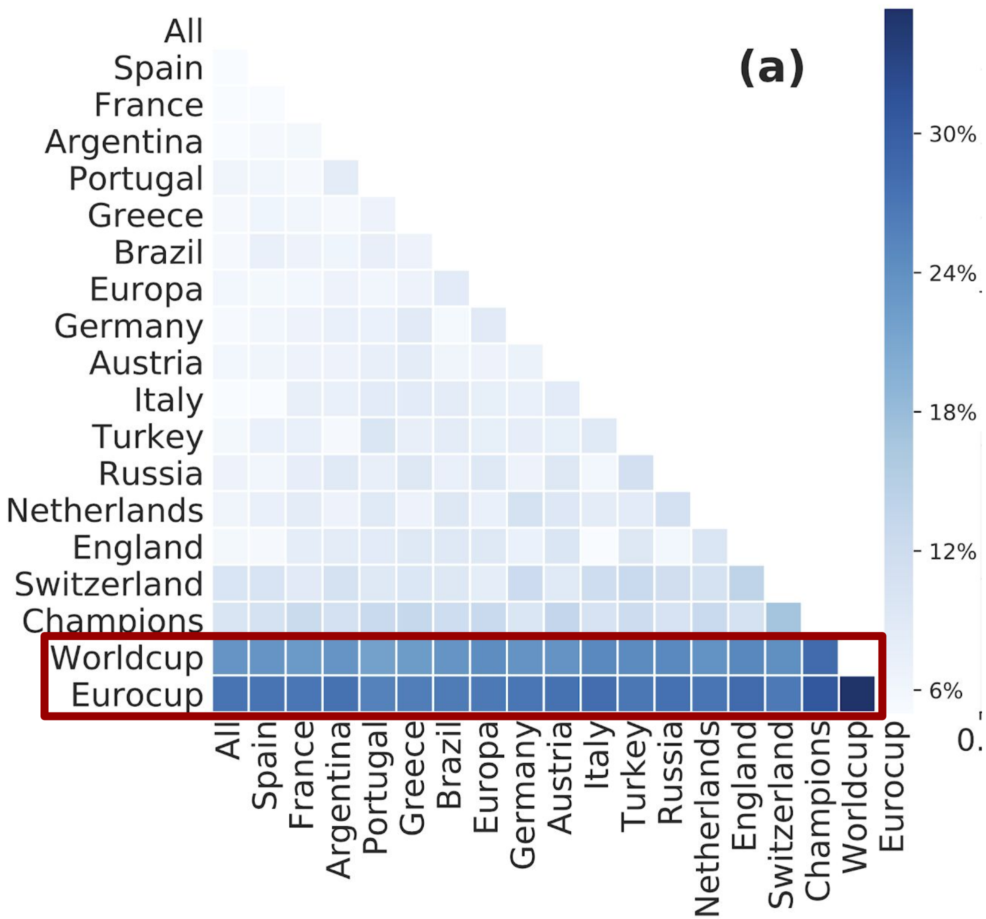
Evaluating the weights

- **stability**
  across competitions and roles
- evaluation of resulting ranking

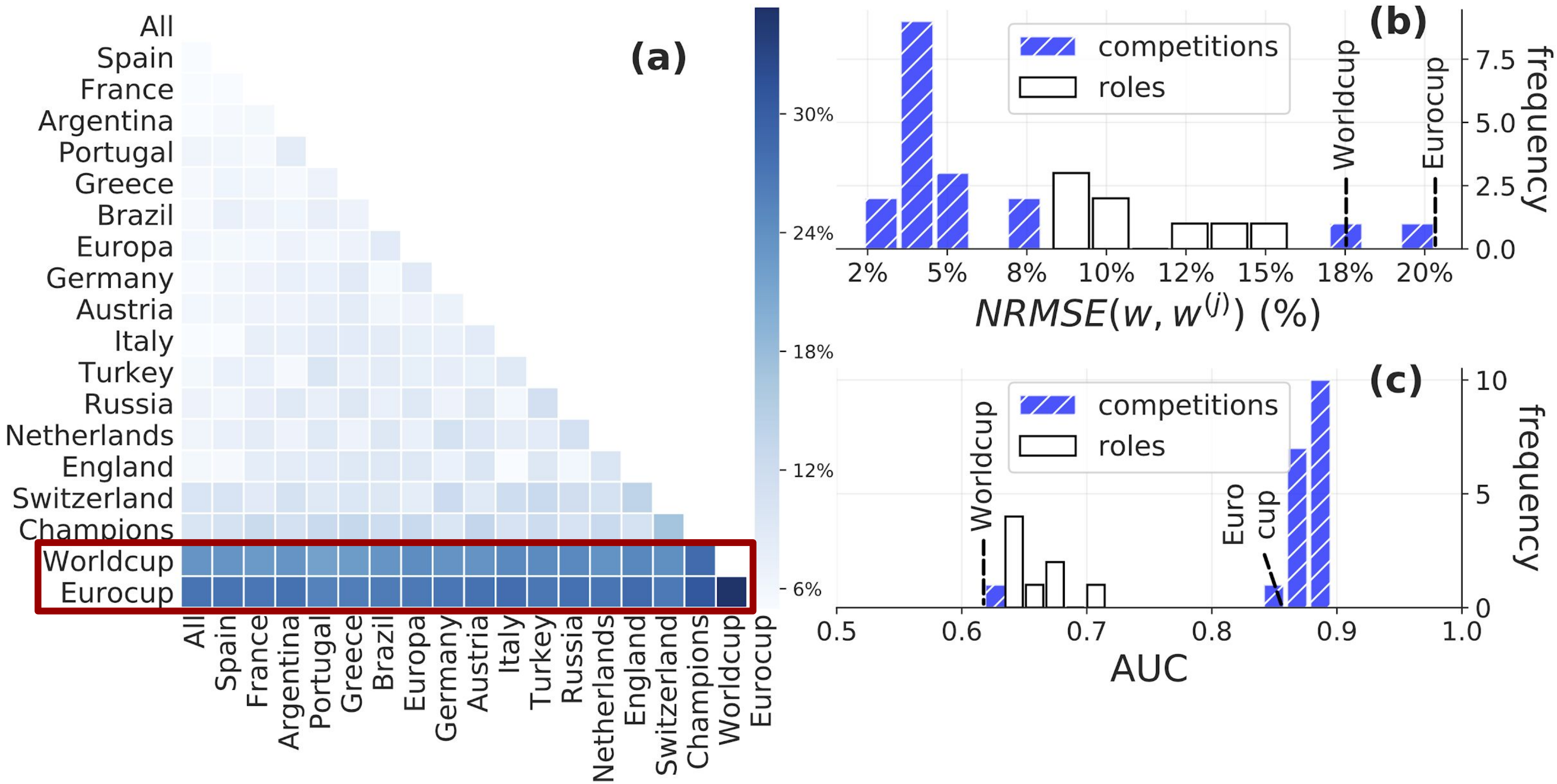
Are these weights “universal”?



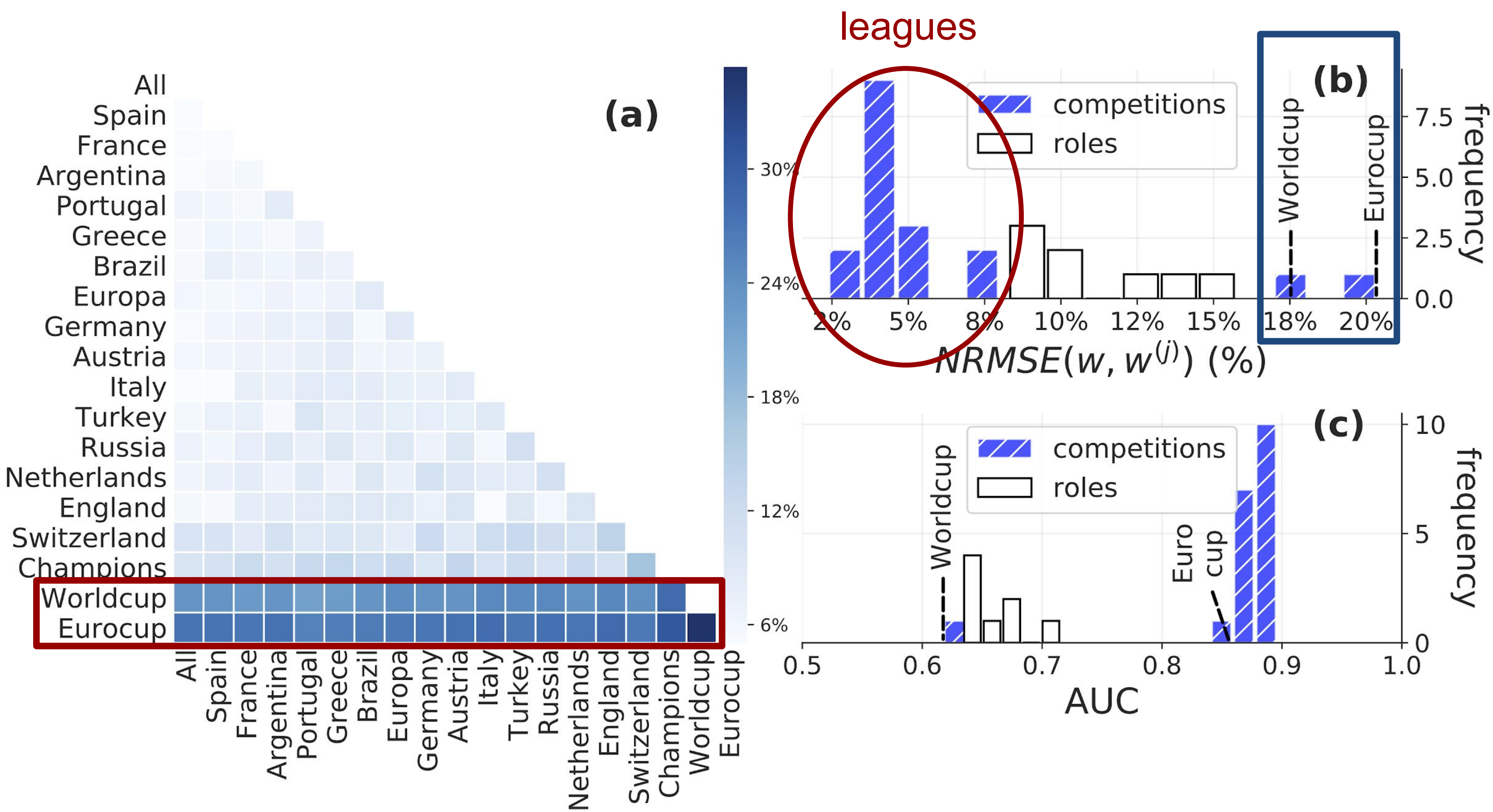
Are these weights “universal”?



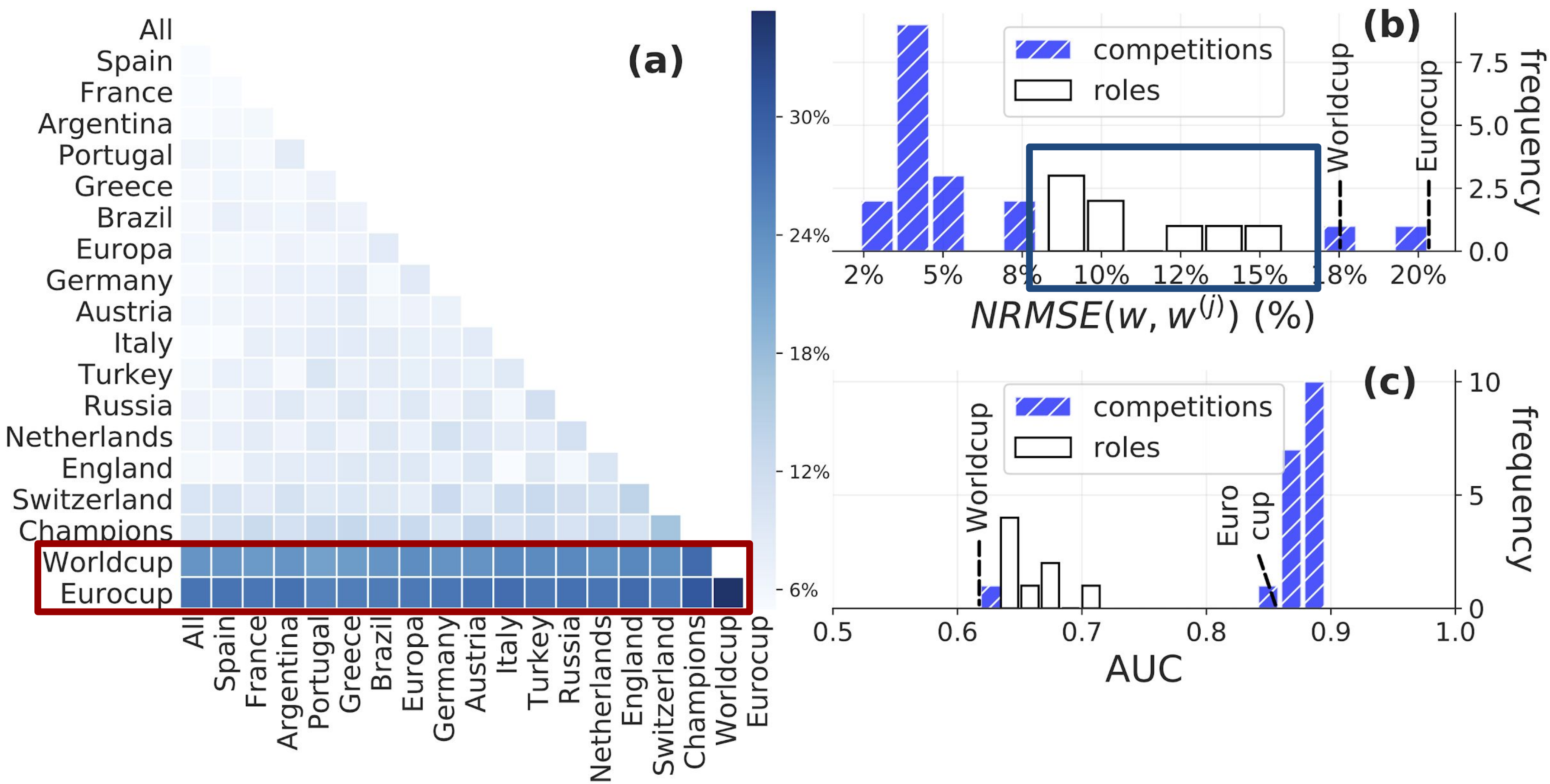
Are these weights “universal”?



Are these weights “universal”?



Are these weights "universal"?



Step #3: role classification

"All ~~animals~~ players are equal, but some ~~animals~~ players are more equal than others."

George Orwell

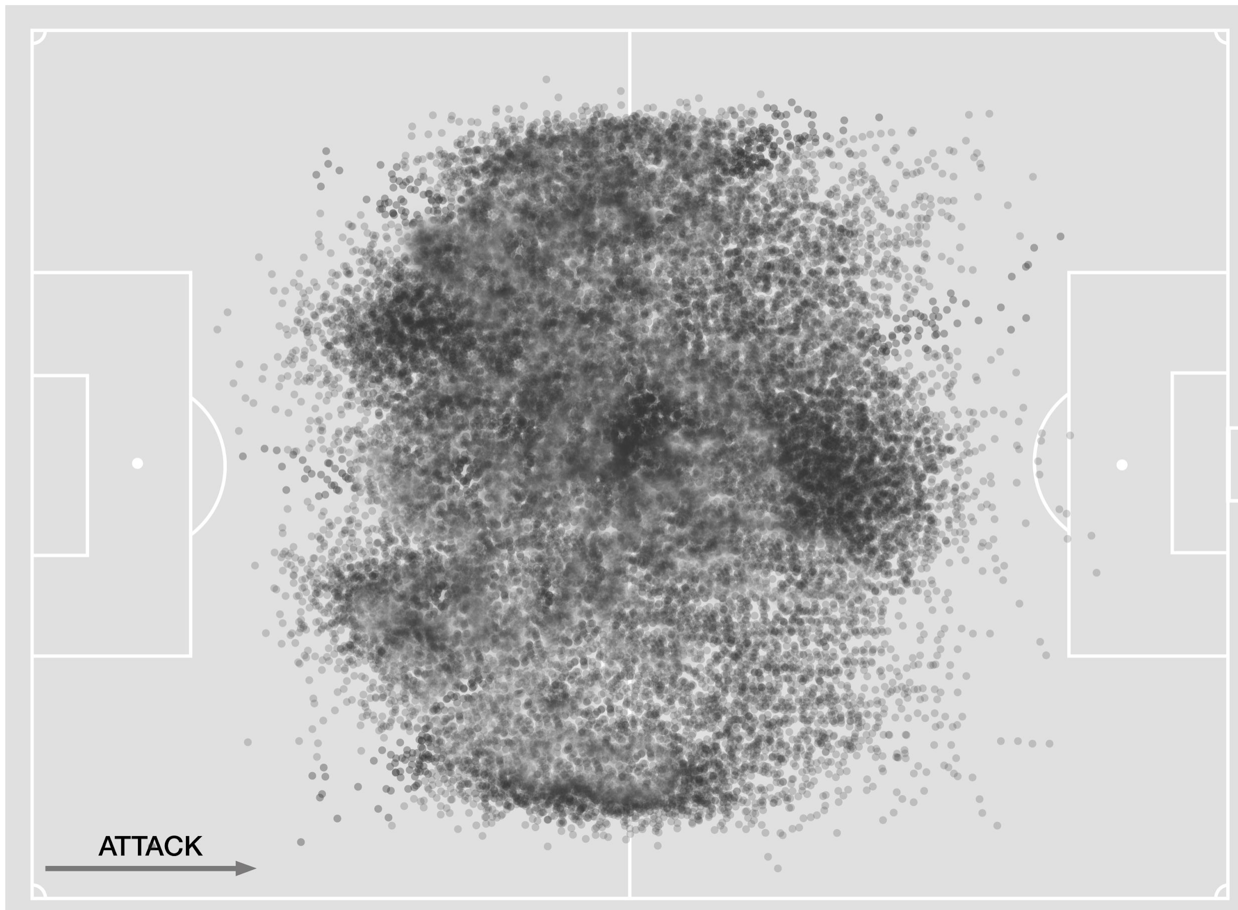
It is meaningless to compare
two players with different roles

Step #3: role classification

```
from pymongo import MongoClient

# load the centers data
res = events.aggregate(pipeline)
X = extract_data(res)
```

```
pipeline = [
    {'$project': {'positions': {'$arrayElemAt': ['$positions', 0]},}},

    {'$group': {
        'x_positions': {'$push': '$positions.x'},
        'y_positions': {'$push': '$positions.y'}
    }},

    {'$project': {
        'avg_x': {'$avg': "$x_positions"},
         'avg_y': {'$avg': "$y_positions"}
     }}]
```

Step #3: role classification

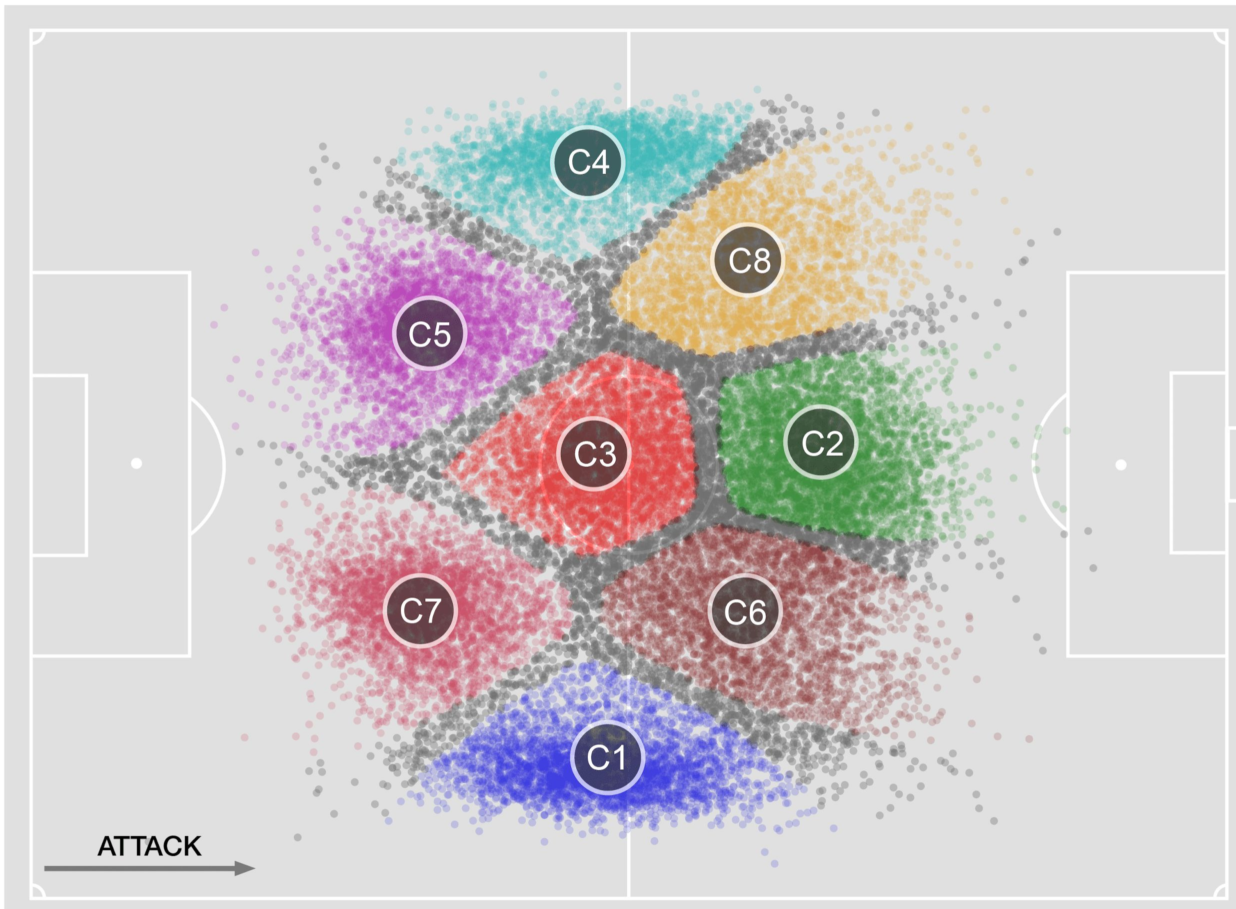


Step #3: role classification

```
from playerank import RoleClusterer

# perform multi-clustering
rc = RoleClusterer(k_range=(2, 20),
   border_threshold=0.2, random_state=42)
rc.fit(X)

rc.labels_
```

```
[[6], [2, 4, 6], [3], [6], [1], [5], [3], [4], [1],
[6], [1], [0], [2, 5], [2], [7], [4], [5], [5], [0],
[4], [5], [4], [6], [3], [5], [1], [6], [4], [0], [7],
[1], [7], [2], [5], [7], [0, 5], ....]
```

Step #3: role classification



Step #4: rating computation

performance rating
of ***u*** in game ***g***

$$r(u,g) = \frac{1}{R}\sum_{i=1}^{n} w_i \times x_i$$

taking into account
the number of goals

$$\alpha \times norm\_goals + (1-\alpha) \times r(u,g)$$

Step #4: rating computation

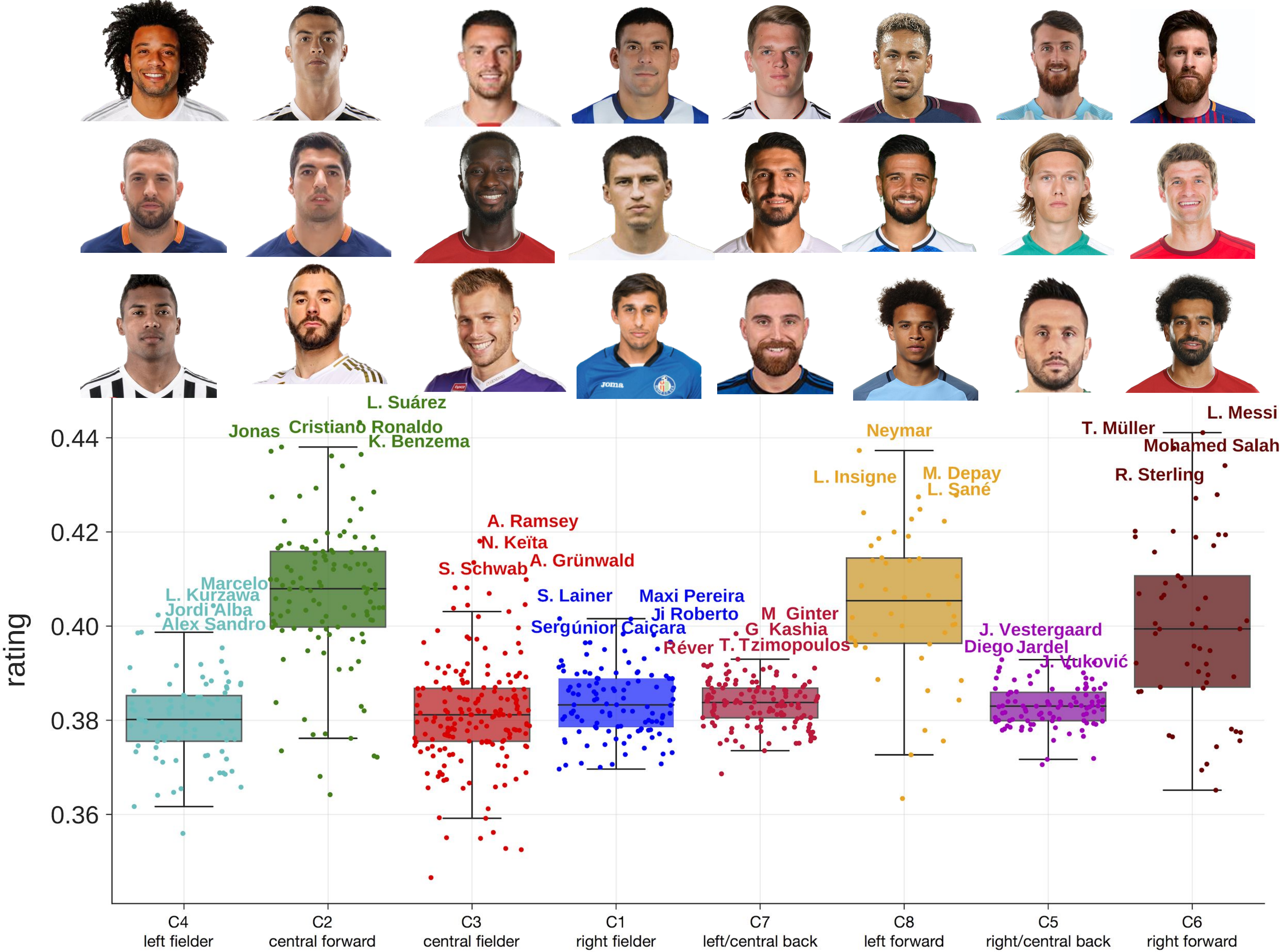
```
from playerank import Rater

res = events.map_reduce(map_aggregate,reduce_sum)
X = extract_data(res)

# rate the performances
rater = Rater(alpha=0.0)
rater.predict(X)
```

goals are not considered

Step #5: player ranking

The ranking of players (by role) can be computed by aggregating over all ratings of the players

```
import pandas as pd

df = pd.read_csv('evaluations.csv')
df.groupby('player_id').mean().sort_values(
              by='rating', ascending=False)
```

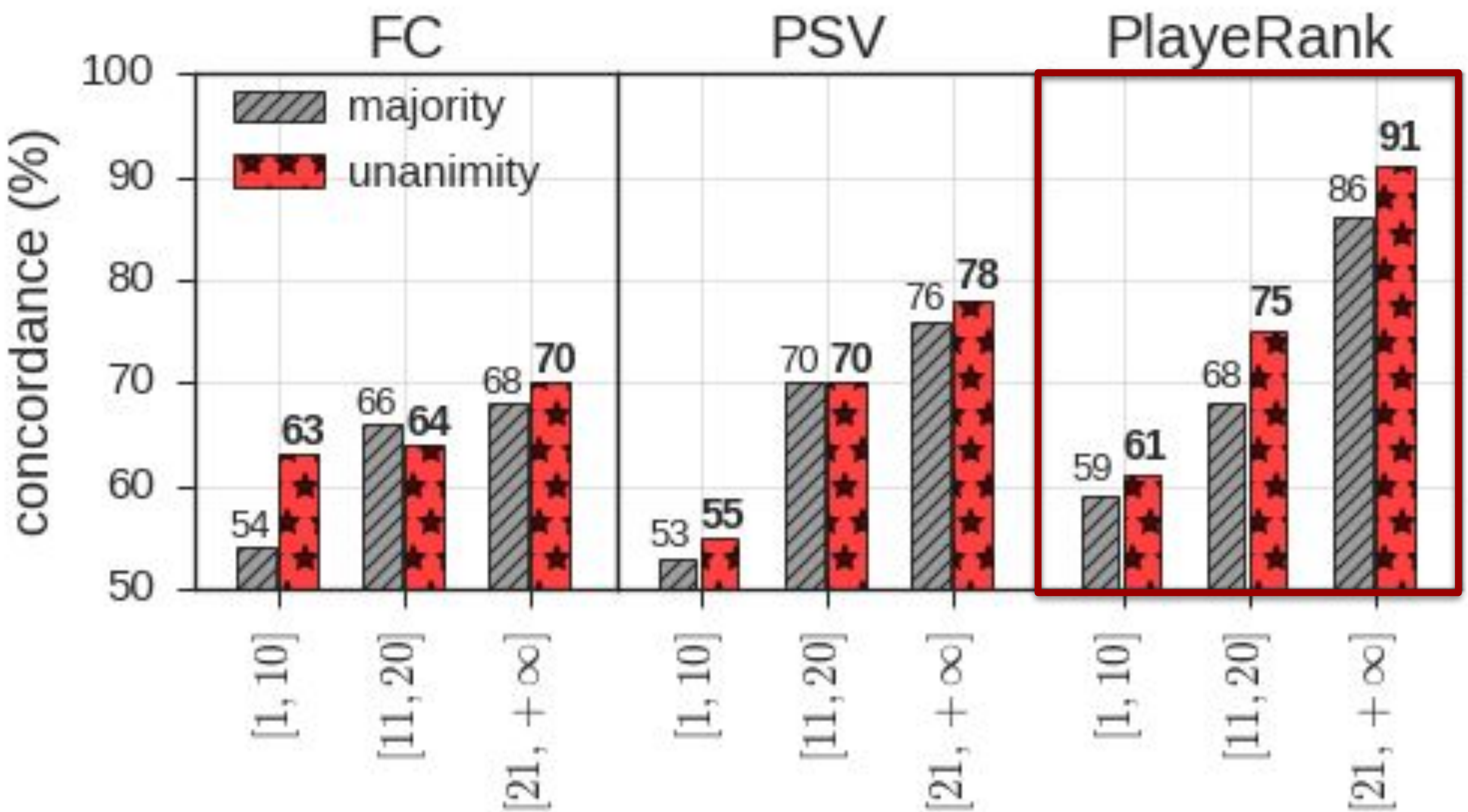

rating
0.44
0.42
0.40
0.38
0.36
L. Suárez
Jonas
Cristiano Ronaldo
K. Benzema
Marcelo
L. Kurzawa
Jordi Alba
Alex Sandro
A. Ramsey
N. Keïta
A. Grünwald
S. Schwab
S. Lainer
Maxi Pereira
Ji Roberto
Sergúnior Caiçara
M. Ginter
G. Kashia
Réver
T. Tzimopoulos
Neymar
L. Insigne
M. Depay
L. Sané
J. Vestergaard
Diego Jardel
J. Vuković
L. Messi
T. Müller
Mohamed Salah
R. Sterling
C4
left fielder
C2
central forward
C3
central fielder
C1
right fielder
C7
left/central back
C8
left forward
C5
right/central back
C6
right forward

How to evaluate the evaluation?

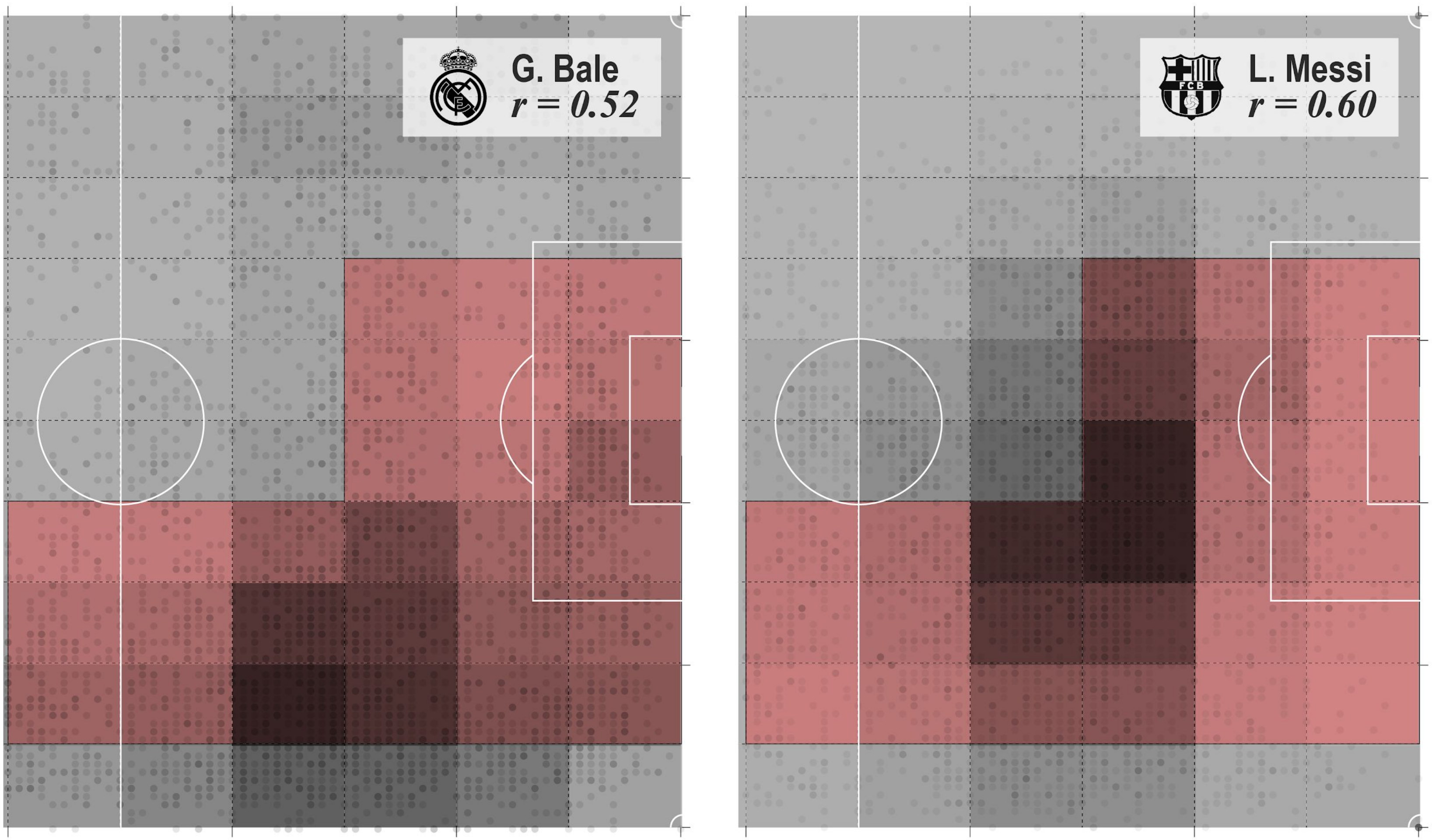
algorithm | expert 1 | expert 2 | expert 3

- majority agreement
- unanimity agreement

Evaluation of 211 pairs



Step #6: the search engine



(a) (b)

Step #6: the search engine

	player	$\hat{r}$	r	$\bar{r}$	club
1	L. Messi	**0.28**	0.60	**0.46**	Barcelona
2	A. Robben	0.26	**0.61**	0.43	Bayern M.
3	L. Suárez	0.24	0.54	0.45	Barcelona
4	T. Müller	0.24	0.56	0.43	Bayern M.
5	Mohamed Salah	0.24	0.56	0.43	Liverpool
6	R. Lukaku	0.24	0.56	0.42	Man. Utd
7	A. Petagna	0.23	0.55	0.42	Atalanta
8	D. Berardi	0.22	0.54	0.41	Sassuolo
9	Aduriz	0.22	0.55	0.40	A. Bilbao
10	G. Bale	0.22	0.52	0.43	R. Madrid

SoBigData

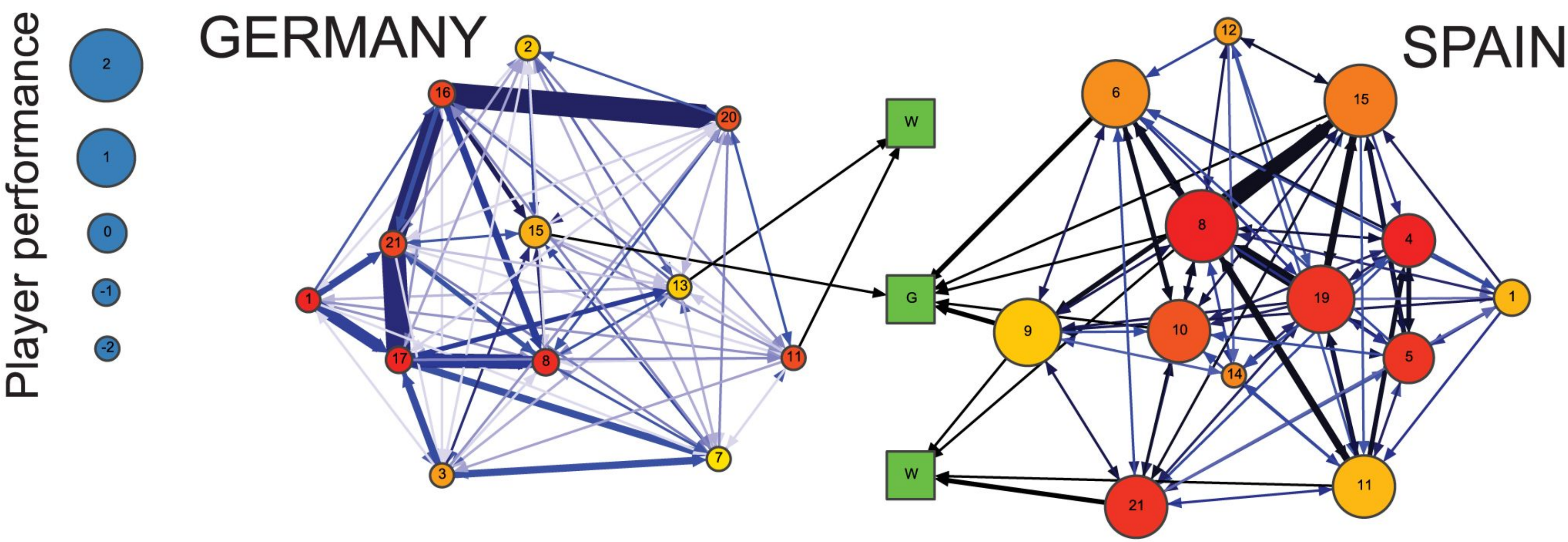
Coming soon:

Soccer Data Challenge @InternetFestival, Pisa, 12-13 October 2018

http://www.internetfestival.it/

Flow Centrality (FC)

Duch et al. (2010) Quantifying the Performance of Individual Players in a Team Activity. PLoS ONE 5(6): e10937.

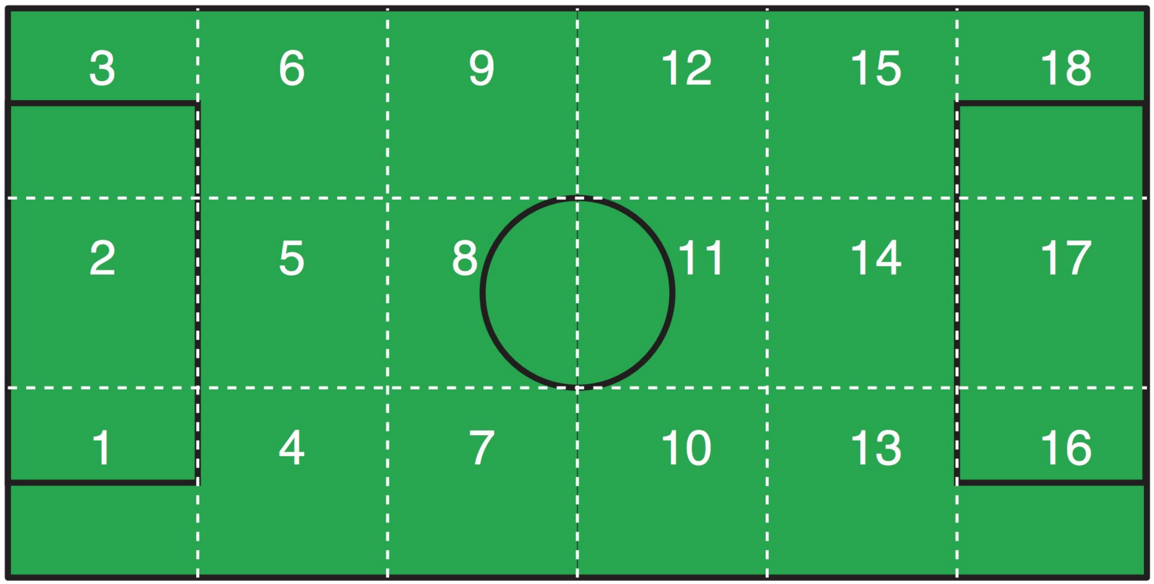


fraction of a player's accurate shots

Validation: 8 of the 20 players in the list of the competition's best players

Pass Shot Value (PSV)

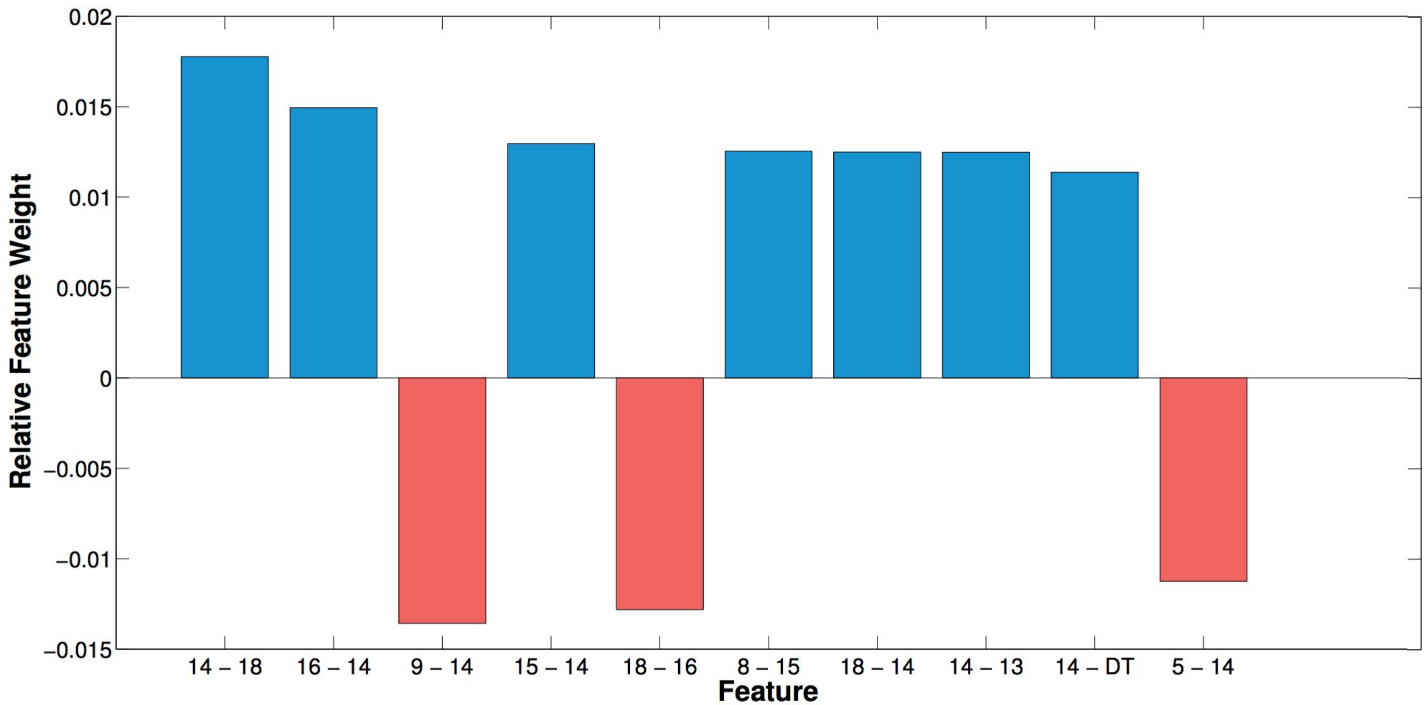
Brooks et al. (2016) Developing a Data-Driven Player Ranking in Soccer using Predictive Model Weights, SIGKDD



each pass is represented as a vector size=360

Pass Shot Value (PSV)

Brooks et al. (2016) Developing a Data-Driven Player Ranking in Soccer using Predictive Model Weights, SIGKDD



predicting if a possession ends in a shot

Validation: correlation with assists and goals