Evaluating passes in premiere league 2017-2018

Karim Elgammal
elgammal@kth.se

***Abstract*—This document is proposing a calculational model for evaluating set of players on their passing efficiency and rank the players according to a calculating scoring metric judging their passes dangerous effictivness**

I. Introduction

Evaluating passes and ranking football players, respectively is an exciting problem to discuss. Some players choose to take the risk and do passes within the last third of the pitch where opponent defenders and goalkeeper are so focused. Thus, building a model in which shall quantify the players' passes be convenient for decision-makers at clubs. With this, I represent a tool that constructs a model for pass evaluation, gives a corresponding score, then rank the players accordingly.

II. Method

A. Analysing Wyscout data and building the model

Wyscout data are processed to construct a logistic regression model. Some of the features in the model:

- passes coordinates, their polynomials
- passes length, passes euclidean distance to opponent GK
- time of the passes

Note: The pass speed can't be a feature as only 10% has valid time-stamps (NaN fields)

B. A walk-through in the code

The model here is tested against the English premiere league. Model construction went through the following steps:

- loading the event data in python
- extracting the passes
- selecting a random subsample of the data (50,000 entries for calculation time consideration)
- building the data in a dataframe according to the features presented in Table III-A

C. Statistical logistic regression model

The features presented in Table III-A are trained in a logistic regression calculation. The resultant 'b-values' are described in Table III-A and arranged in descending order.

D. Judging passing events and recording their score

The model is giving the players a score in order to rank the players according to these steps:

- get all the passes done by the player using 'playerId'
- get all the continuous score predictions using the model for these passes
- aggregate these predictions using mean
- compare results from different model experiments to fix over-fitting description by measuring test and train accuracy
- normalizing the data 0-1 to improve gradient descent and convergence

E. Multilayer perceptron (for better accuracy)

Tested the model against a neural network of 100 hidden layers and improved the model test and train accuracy by 1%.

F. Testing the model with dangerous passes hypothesis

Model 2 takes consideration of the passes happening at the last third, closer to the opponent goal keeper. But the model accuracy is pretty suppressed as indicated in Table IV.

G. Calculations models

There are two models differentiated according to Table IV

III. Results

Results regarding the effective passes that the models computed are compiled here.

A. Coefficients

Table III-A shows the model coefficients in descending order. Some important features affecting the passing model are the starting distance of the pass to the opponent GK, c-coordinate and coordinate powers as well as when during the game the pass occurred.

TABLE I
Regression coefficients values and model features

Features (coefficients)	coefficient values	no. in code
Dist. to keep start	(+) 5.524292676208213	7
c start squared	(+) 4.808404582425619	13
c end	(+) 3.529119006804113	6
X∗Y start squared	(+) 2.2581505439145158	17
d Distance	(+) 2.5255947184507415)	20
Y end squared	(-) 2.3480386584756294	15
match period	(-) 1.9929521807011261	10
x start	(-) 1.6975419531263858	0
Y end	(+) 1.4719043681376969)	5
dX	(-) 1.279736376058593)	19
d distance	(-) 1.1906924267726056	18
Y start squared	(-) 0.9349654384868694	12
c end squared	(-) 0.9316298605030615	16
X end squared	(-) 0.6808026691089022)	14
Distance to keep end	(-) 0.6367630749018166)	8
C start	(+) 0.5986214246264059	3
X start	(+) 0.24872950006684913	1
X start squared	(+) 0.03160717107523749	11
Distance pass	(+) 0.005244728727661594	9
Y start	(-) 0.003805495437266778	2

B. Model resultant best and worst players

The model can list the top ten and worst ten players and their respective efficiency score, illustrated in the method section. Model 1 players are shown in Tables IV, IV respectively. Similarly, Model 2 results are shown in Tables IV, IV. Model 2 is differentiated from model 1 by restricting the successful passes at the last third of the pitch closer to the opponent GK. Meanwhile, model 2 accuracy is pretty low as indicated in Table IV, needs further refine.

IV. Discussion

It appears that model 1 had found some real good passers up-front like Ibra and Afellay, yet, it failed into the trap of (maybe) some safe passers represented few resultant GK, CB. Note that age-limit is not constrained. By only including the last third (closer to the opponent GK) successful passing events, model 2 found some excellent example players that usually have a role either in attacking or counter attack as Phil Foden, Diaz whom both were pretty young at that time and showed good potential despite their limited played minutes. Other good examples I believe are both Barkley and Jones, who both usually perform high-pressure at the end of the game when Chelsea and Man Utd need to get the match points. I can't deny that model 2 can fall into the curse of 'time-wasting' short passes close to the corner area. Meanwhile, the resultant worst players, according to model 2 are mostly GKs which is expected (Table IV). Redoing model 2 on doubled sample events reveals some overlap and confirms that Man City academy raised some good passers in-front.

For further investigations, I believe that the time where the pass is taking place during the game has somehow an effect on the threat it can do especially with higher fatigue.

TABLE II
CALCULATIONAL MODELS DETAILS

model #	training accuracy	testing accuracy	samples	condition
1	85%	84%	50000	
2	72%	73%	50000	last 3rd pass

TABLE III
MODEL 1 TOP 10 PLAYERS

	player name	score	position	club
1	Uwe Hünemeier	0.97	CB	Brighton & Hove Albion
2	Jay Fulton	0.96	MD	Wigan Athletic
3	Muhamed Bešić	0.94	DM	Middlesbrough
4	Joe Willock	0.94	MD	Arsenal
5	Ethan Ampadu	0.93	CB	Chelsea
6	Paulo Gazzaniga	0.93	GK	Southampton
7	Michael Hefele	0.92	CB	Huddersfield Town
8	Zlata Ibrahimovic	0.92	ST	Man Utd
9	Stephen Ireland	0.92	CAM	Stoke City
10	Ibrahim Afellay	0.92	CAM	Stoke City

V. Acknowledgment

Ack. team one for their discussion and help specially to Christian Gilson and Caglar Altunel sharing code snippets.

TABLE IV
MODEL 2 TOP 10 PLAYERS

	player name	score	position	club
1	Phil Foden	0.87	LM	Man City
2	Brahim Diaz	0.86	FW	Man City
3	Kostas Stafylidis	0.84	LB	Fulham
4	Cenk Tosun	0.83	ST	Everton
5	Ross Barkley	0.82	CAM	Chelsea
6	Per Mertesacker	0.82	CB	Arsenal
7	Phil Jones	0.82	CB	Man Utd
8	N. Wells	0.82	FW	Burnley
9	S. Kaikai	0.81	WF	Crystal Palace
10	I. Afellay	0.81	CAM	Stoke City

TABLE V
MODEL 1 WORST 10 PLAYERS

	player name	score	position	club
1	Levi Lumeka	0.32	WF	Crystal Palace
2	Eldin Jakupovic	0.53	GK	Hull City
3	Tyrese Campbell	0.59	CF	Stoke City
4	Nick Pope	0.60	GK	Burnley
5	Nordin Amrabat	0.61	WF	Watford
6	Jack Butland	0.65	GK	Stoke City
7	Orestis Karnezis	0.64	GK	Watford
8	Alex McCarthy	0.64	GK	Southampton
9	Jordan Pickford	0.64	GK	Everton
10	Jérémy Pied	0.62	RB	Southampton

TABLE VI
MODEL 2 WORST 10 PLAYERS

	player name	score	position	club
1	Joe Hart	0.19	GK	West Ham Utd
2	Eldin Jakupovic	0.20	GK	Leicester City
3	Fraser Forster	0.24	GK	Southampton
4	Jacob Murphy	0.25	WF	Newcastle United
5	Ederson	0.25	GK	Man City
6	Lee Grant	0.26	GK	Stoke City
7	Younès Kaboul	0.31	CB	Watford
8	Łukasz Fabiański	0.31	GK	West Ham Utd
9	Hugo Lloris	0.31	GK	Tottenham Hotspur
10,	Mathew Ryan	0.32	GK	Brighton & Hove Albion

TABLE VII
MODEL 2 BEST 10 (100,000 EVENT SAMPLES TO DOUBLE CHECK)

	player name	score	position	club
1	L. Nmecha	0.89	FW	Man City
2	Brahim Diaz	0.88	FW	Man City
3	P. Mertesacker	0.84	CB	Arsenal
4	Jürgen Locadia	0.84	ST	Brighton & Hove Albion
5	Sullay Kaikai	0.83	WF	Crystal Palace
6	Nahki Wells	0.82	FW	Crystal Palace
7	Abdelhamid Sabiri	0.82	MD	Huddersfield Town
8	Sam Baldock	0.82	ST	Brighton & Hove Albion
9	Cenk Tosun	0.81	ST	Everton
10	Michy Batshuayi	0.81	ST	Chelsea