Prediction of Football Games Results

Roman Nestoruk, Sollers Consulting Sp. z o.o., ul. Koszykowa 54, Warsaw 00-675, Poland

Grzegorz Słowiński, University of Technology and Economics, Engineering Department, ul. Jagiellońska 82f, Warsaw 03-301, Poland

Agenda

- Motivation

- Problems

- Structure of model

- Prediction example

  - Example of data

  - Validation

- Future development

Motivation

Football statistic is the source of unlimited data with different quality as over the world, more than 100 official football games are played every day and a lot of different factors would impact each of those games, players, tactics, cards, injuries and etc.

Questions:

Can we predict future games results?

What data needed?

How efficient we can do this?

Problems

- As there is no data in correct format accessible by downloading or API connection, all of them should be collected manually

- Data for the current game, are correct only until the tame, the game has started

- Bad data quality

- Some unpredictable games event (fans running to the field, injuries)

- Any two games will have their unique scenario

Example of input data

Example is following for the game Celta Vigo vs Granada

Page with statistics looks following:

Tournament	Apps	Goals	Shots pg	Discipline		Possession%	Pass%	AerialsWon	Rating
LaLiga	7	5	8.1	28	0	38.1	68.7	16.6	6.43
Copa del Rey	5	15	N/A	5	0	N/A	N/A	N/A	-
Club Friendlies	5	6	N/A	0	0	N/A	N/A	N/A	-
Total / Average	17	26	8.1	33	0	38.1	68.7	16.6	6.43

Based on this statistic, we can have data in format:

Home team (Celta Vigo): 6.68,0,3,0,0,0,1,2,2,3,0,0,3,1.1,52.1,81.7,10.3,7,6.6,6.5,6.5,6.9,7.5,6.7,6.9,6.8,7,7.2

Away team (Granada): 6.87,2,0,2,0,0,2,1,0,2,1,2,0,1.4,55.6,81.7,10.3,6.8,6.7,7.1,6.9,7,6.8,7,6.6,7,7,6.2,6.5

Structure of model

Hidden layer 1 Consist of 56 units, with RELU activation.

Hidden layer 2 Consist of 28 units, with RELU activation.

Output layer Consist just of 2 units, with Linear activation.

```
Layer (type)                 Output Shape              Param #
=================================================================
dense_9 (Dense)              (None, 56)                3192
_____
dense_10 (Dense)             (None, 28)                1596
_____
dense_11 (Dense)             (None, 2)                 58
=================================================================
Total params: 4,846
Trainable params: 4,846
Non-trainable params: 0
```

CS&P'21

Validation

For this example, we will you data from 27.09.2021 games:

Crystal Palace vs Brighton

Celta Vigo vs Granada

Venezia vs Torino

Crystal Palace	Draw	Brighton
3,05	3,05	2,60
Exact scores:		
0-1	6,8	
0-2	10,50	
Celta Vigo	Draw	Granada
1,66	3,95	5,60
Exact scores:		
1-0	6,9	
1-1	7,20	
Venezia	Draw	Torino
3,45	3,25	2,26
Exact scores:		
0-0	7,90	
0-1	7,40	

Validation

Teams		Model 1		Model 2		Model 3		Agenda:	
Crystal Palace	Brighton	0,02129224	0,16984874	0,016044421	0,10254879	0,041102193	0,14092396	Expected to win	
1	1	0,2129224	1,6984874	0,16044421	1,0254879	0,41102193	1,4092396	Expected to lose	
		0	2	0	1	0	1	Expected to drowe	
Celta Vigo	Granada	0,08358175	0,027207743	-0,09241221	-0,057339683	0,06111416	0,08069539	Predicted corrected result	
1	0	0,8358175	0,27207743	-0,9241221	-0,57339683	0,6111416	0,8069539	Predicted exact score	
		1	0	0	0	1	1		
Venezia	Torino	0,008526646	0,043268904	0,020941237	0,031503078	0,003891431	0,13276331		
1	1	0,08526646	0,43268904	0,20941237	0,31503078	0,038914313	1,3276331		
		0	0	0	0	0	1		
Models evaluation for 27.09.2021		Similar to odds:	2,00	Similar to odds:	1,00	Similar to odds:	2,00		
		Predicted:	2,00	Predicted:	1,00	Predicted:	0,00		
		Exact score:	1	Exact score:	0	Exact score:	0		
		Win:	11,81	Win:	3,25	Win:	0		
		Investment:	6	Investment:	6	Investment:	6		
		Benefit:	5,81	Benefit:	-2,75	Benefit:	-6		

Future development scenario

- Getting more data with better quality

- Updating model's structure

- Appling data from second categories tournaments

Extra example

England	Italy	0,20153251	0,22888476	0,10953673	0,08836879	0,18839952	0,1299682
1	1	2,0153251	2,2888476	1,0953673	0,8836879	1,8839952	1,299682
		2	2	1	1	2	1

Incoming data:

England: 6.84;1;0;0;0;1;0;2;0;4;0;2;1;1.79;57.8;87.7;12.8;6.6;6.8;6.9;7.7;7.5;6.7;6.7;6.5;7.4;7.6;7.8

Italy: 6.90;3;0;3;0;1;0;2;1;2;1;1;1;2.42;56.5;86.5;17.9;6.9;7;6.9;6.8;7.3;6.8;7.4;7.6;6.9;6.8;7.2

Outcome of 3 models: 0.20153251-0.22888476;0.10953673-0.08836879;0.18839952-0.1299682

Thank you!

Q/A