World Cup Team Evaluation Model based on TOPSIS Method through FIFA Player Data Analysis

Chang Shi

Oregon State University, Corvallis, 97330, United States

Abstract. The purpose of this study is to analyze which teams are stronger and more likely to win the football World Cup. Although there are many classic cases of the weak winning the strong because of the obvious differences between the national football teams, these are only a very small part and will be excluded from consideration in this article. This article will analyze the performance scores of players in their clubs given by FIFA through the TOPSIS method, and finally the ranking of national football teams. As the current world football rankings are based on points accumulated from wins and losses in intercontinental competitions and friendlies, they do not give a specific indication of a team's strength. This article will rate the highest ranked teams in each group stage of the World Cup and finally rank them. Finally, entropy weight method and simple TOPSIS method are analyzed which is more suitable for the analysis of football World Cup.

Keywords: Football World Cup; FIFA Data; TOPSIS; Entropy Weight Method.

1. Introduction

Football is one of the most popular sports in the world, with 4 billion fans (Jason Shvili). The last football World Cup, the 2018 World Cup in Russia, was watched by 3.572 billion viewers and the final was watched by 1.1 billion people (FIFA.com). The 2022 World Cup in Qatar will be held in November, and even though the tournament hasn't started yet, the tournament is already getting a lot of attention. People in most countries hope that their country can participate in this competition and hope that their country can win the championship. However, only 32 teams can participate in this World Cup and only one team can win the final championship. World Cup matches attract countless people to win the championship, even in the final of the last World Cup in Russia, the top leaders of two countries, France, and Croatia, arrived to watch the game. It is a great honor for every player.

There are eight groups in the Football World Cup, and each group has four teams. This article will analyze the eight groups from the perspective of ranking the first, and then rank these eight teams. The data are based on the current official FIFA team rankings and the combined score given by FIFA for a player's club performance. The top nine teams in the world are Brazil, Belgium, Argentina, France, England, Spain, Italy, the Netherlands, and Portugal, but Italy has no way to qualify for the World Cup, so Italy has been left out (FIFA.com).

In addition, the method used in this article is TOPSIS and entropy weight method. The football team has 11 players on the pitch and the World Cup squad has 23, so this article will evaluate the strikers, Super League players, defenders, and goalkeepers to give an overall rating.

2. Data Analysis with TOPSIS (body)

TOPSIS, known as technique for Order of Preference by Similarity to Ideal Solution, is a multi-criteria decision analysis method. It compares a set of alternatives against pre-specified criteria. This approach is used in businesses across a wide range of industries, each time we need to make analytical decisions based on collected data (Robert Soczewica). This paper will use TOPSIS to score and rank 8 teams, then finally analyze whether TOPSIS is suitable for the field of football.

The preprocessing part of the data is to extract the total rating, country, name, and position of the player for FIFA2022(STEFANO LEONE). Since the national football team consists of 23 players, most of which are made up of six forwards, six midfielders, eight defenders and three goalkeepers, so what the data will show is Brazil, Belgium, Argentina, France, England, Spain, the Netherlands,

and Portugal with 6 forwards, 6 midfielders, eight defenders and three goalkeepers are the highest rated players.

Table 1. Average players in Brazil, Belgium, Argentina and France

Brazil	Overall	Belgium		Argentina		France	
W/F	84	W/F	83.83	W/F	87.33	W/F	86.5
M	86.67	M	82.83	M	84	M	85.3
B	84	B	78.88	B	79.63	B	83.5
GK	86.67	GK	85	GK	82	GK	84.3

Table 2. Average of players in England, Spain, Holland and Portugal

England		Spain		Netherlands		Portugal	
W/F	86.17	W/F	84.67	W/F	81.33	W/F	84.5
M	83.67	M	86	M	83.5	M	84.3
B	84	B	84.25	B	83.75	B	81.75
GK	82	GK	83	GK	78.33	GK	81.7

Step 1: After that, the data is processed symbolically, X1 = Brazil, X2 = Belgium, X3 = Argentina, X4 = France, X5 = England, X6 = Spain, X7 = the Netherlands, X8 = Portugal, A1 = W/F, A2 = M, A3 = B, A4 = GK, then the table 3 is

Table 3. Matrix of data

	X1	X2	X3	X4	X5	X6	X7	X8
A1	84	83.83	87.33	86.5	86.17	84.67	81.33	84.5
A2	86.67	82.83	84	85.33	83.67	86	83.5	84.33
A3	84	78.88	79.63	83.5	84	84.25	83.75	81.75
A4	86.67	85	82	84.33	82	83	78.33	81.67

$$\alpha_{ij} = \frac{a_{ij}}{\sqrt{\sum_{i=1}^{M} {a_{ij}}^2}}$$

Table 4. Normalized data

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.492	0.507	0.524	0.509	0.513	0.501	0.497	0.509
A2	0.508	0.501	0.504	0.502	0.498	0.509	0.511	0.508
A3	0.492	0.477	0.478	0.492	0.500	0.499	0.512	0.492
A4	0.508	0.514	0.492	0.497	0.488	0.491	0.479	0.492

Step 2: Now let's not consider the weights, let's make each factor equally important, then ω_j is $\frac{1}{4}$.

$$X_{ij} = \alpha_{ij} \times \omega_j$$

$$\omega_j = \frac{\omega_j}{\sum_{j=1}^{N} \omega_j}$$

$$\sum_{j=1}^{N} \omega_j = 1$$

Table 5. Perform 1:1:1:1 after the weight data

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.123	0.127	0.131	0.127	0.128	0.125	0.124	0.127
A2	0.127	0.125	0.126	0.126	0.125	0.127	0.128	0.127
A3	0.123	0.119	0.120	0.123	0.125	0.125	0.128	0.123
A4	0.127	0.129	0.123	0.124	0.122	0.123	0.120	0.123

Step 3: Distinguish the maximum and minimum values of data

$$X_j^b = max_{i=1}^M X_{ij}$$
$$X_j^\omega = min_{i=1}^M X_{ij}$$

Sixteen values will appear:

Table 6. Sixteen values

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.123	0.127	0.131	0.127	0.128	0.125	0.124	0.127
A2	0.127	0.125	0.126	0.126	0.125	0.127	0.128	0.127
A3	0.123	0.119	0.120	0.123	0.125	0.125	0.128	0.123
A4	0.127	0.129	0.123	0.124	0.122	0.123	0.120	0.123

Step 4: Calculate the Euclidean distance between the target alternative and the best/worst alternative: (Robert Soczewica)

$$d_i^b = \sqrt{\sum_{j=1}^{N} (X_{ij} - X_j^b)^2}$$
$$d_i^\omega = \sqrt{\sum_{j=1}^{N} (X_{ij} - X_j^\omega)^2}$$

Table 7. Calculate the Euclidean distance between the target alternative and the best/worst alternative

	X1	X2	X3	X4	X5	X6	X7	X8
d b	0.00552	0.00998	0.0149	0.00571	0.00792	0.00550	0.00908	0.00595
d w	0.00552	0.0133	0.0138	0.00531	0.00731	0.00540	0.0123	0.00586

Step 5: For each alternative calculate the similarity to the worst alternative. The results are our TOPSIS scores. (Robert Soczewica)

$$S_i = \frac{d_i^\omega}{d_i^\omega + d_i^b}$$

Table 8. TOPSIS scores

	X1	X2	X3	X4	X5	X6	X7	X8
Si	0.5	0.572	0.480	0.482	0.480	0.495	0.576	0.496

Now making a ranking, according to the higher the score, the higher the ranking principle.
S7 > S2 > S1 > S8 > S6 > S4 > S5 >S3.

From TOSIS, it is feasible to draw the conclusion that the Dutch football team ranked first, the Belgian football team in the second, the Brazilian football team in the third, fourth, Portugal but that does not conform to the real conditions, because Brazil is the first international world rankings, Belgium is the second, so this article will discuss a new method, the entropy weight method, to analyze this issue.

3. Mathematical Model Entropy Weight Method (body)

The entropy weight method based on the evolution of TOPSIS is basically the same as TOPSIS, except that the second part of the weight of TOPSIS is established. The weighting ratio of TOPSIS used in this paper is established in the case of 1:1:1:1, but it is impossible to have this situation in practice. In football, there is always a certain part of the team that is the most important, and many managers plan their tactics around it. Therefore, a scientific and technical proportion of the entropy weight method will be more appropriate (Yuxin Zhu). This will be an improvement on TOPSIS, and the entropy weight method based on the evolution of TOPSIS will be more scientific and digital.

Step 1: The data normalized by TOPSIS are used here.

Table 9. The data normalized by TOPSIS are used

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.492	0.507	0.524	0.509	0.513	0.50106453	0.49741232	0.50858555
A2	0.508	0.501	0.504	0.502	0.498	0.50895531	0.51066306	0.50758242
A3	0.492	0.477	0.478	0.492	0.500	0.49859866	0.51219199	0.49203395
A4	0.508	0.514	0.492	0.497	0.488	0.49120106	0.47906514	0.49153238

Step 2: Standardization of measurements

$$p_{ij} = \frac{X_{ij}}{\sum_{i=1}^{n} x_{ij}}$$

Table 10. Standardization of measurements

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.246	0.253	0.262	0.255	0.257	0.251	0.249	0.254
A2	0.254	0.251	0.252	0.251	0.249	0.255	0.255	0.254
A3	0.246	0.239	0.239	0.246	0.250	0.249	0.256	0.246
A4	0.254	0.257	0.246	0.248	0.244	0.246	0.240	0.246

Step 3: In EWM, the entropy value Ei of the ith index is defined as (Yuxin Zhu)

$$e_j = -k \sum_{i=1}^{n} p_{ij} \times \ln(p_{ij})$$

$$k = \frac{1}{\ln(n)}$$

ej	1.34	1.34	1.33	1.33

Step 4: The larger Ei is, the greater the divergence of index i is, and the more information can be derived. Therefore, the index should be given higher weight. (Yuxin Zhu)

$$\omega_j = \frac{1 - e_j}{\sum_{j=1}^{m}(1 - e_j)}$$

wj	0.254	0.253	0.246	0.247

Step 5: Now, back to the third step in the TOPSIS step3. W1 is 0.254, w2 is 0.253, w3 is 0.246 and w4 is 0.247.

Table 11. The calculation results

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.125	0.129	0.133	0.129	0.130	0.127	0.126	0.129
A2	0.128	0.127	0.128	0.127	0.126	0.129	0.129	0.128
A3	0.121	0.117	0.118	0.121	0.123	0.123	0.126	0.121
A4	0.126	0.127	0.122	0.123	0.121	0.121	0.118	0.122

Step 6: Calculate the Euclidean distance between the target alternative and the best/worst alternative: (Robert Soczewica)

$$X_j^b = max_{i=1}^{M} X_{ij}$$
$$X_j^\omega = min_{i=1}^{M} X_{ij}$$

Sixteen values will appear:

Table 12. Sixteen values

	X1	X2	X3	X4	X5	X6	X7	X8
A1	0.125	0.129	0.133	0.129	0.130	0.127	0.126	0.129
A2	0.128	0.127	0.128	0.127	0.126	0.129	0.129	0.128
A3	0.121	0.117	0.118	0.121	0.123	0.123	0.126	0.121
A4	0.126	0.127	0.122	0.123	0.121	0.121	0.118	0.122

$$d_i^b = \sqrt{\sum_{j=1}^{N}(X_{ij} - X_j^b)^2}$$
$$d_i^\omega = \sqrt{\sum_{j=1}^{N}(X_{ij} - X_j^\omega)^2}$$

d b	0.00876	0.0116	0.0199	0.0107	0.0126	0.00969	0.0116	0.0110
d w	0.00952	0.0177	0.01890	0.0106	0.0111	0.00931	0.0152	0.0110

Step 7: For each alternative calculate the similarity to the worst alternative. The results are our TOPSIS scores. (Robert Soczewica)

$$S_i = \frac{d_i^\omega}{d_i^\omega + d_i^b}$$

Si	0.521	0.605	0.487	0.496	0.468	0.490	0.5671	0.499

Now making a ranking, according to the higher the score, the higher the ranking principle.

S2 > S7 > S1 > S8 > S4 > S6 > S3 >S5.

From the result of the entropy weight method, Belgium may have a better World Cup performance.

4. Discussion/Conclusion

Having analyzed the performance of FIFA players in their clubs through TOPSIS, this is an analysis method that can provide valid opinions. TOPSIS based on entropy weight method change differs from TOPSIS in the control of weights. The weights changed by entropy weight method can also affect the result. The current ranking of world national teams is based on the number of points earned by the national team during matches. The teams in the European region are generally better than those in the Americas, and at the continental level Brazil is more likely to win, so Belgium has a good chance of winning the final World Cup.

References

[1] FIFA.com. Sep 23, 2022, from https://www.fifa.com/fifa-world-ranking/men?dateId=id13603.

[2] Jason Shvili. The Most Popular Sports in The World. Oct 16, 2020, from https:// www.worldatlas. com/ articles/what-are-the-most-popular-sports-in-the-world.html#: ~:text=1%20Football% 20% 28 Soccer% 29%20is%20the%20most%20popular%20sport, wear%20 much%20 more% 20 protective% 20 gear% 20than%20rugby%20players.

[3] FIFA.com. More than half the world watched record-breaking 2018 World Cup. Dec 21, 2018, from https://www.fifa.com/tournaments/mens/worldcup/2018russia/media-releases/more-than-half-the-world-watched-record-breaking-2018-world-cup.

[4] STEFANO LEONE. FIFA 22 complete player dataset. 2021, from https://www. kaggle.com/ datasets/ stefanoleone992/fifa-22-complete-player-dataset?resource=download.

[5] Robert Soczewica. What is TOPSIS. Sep 26, 2020, from https://robertsoczewica.medium.com/what-is-topsis-b05c50b3cd05.

[6] Yuxin Zhu. Effectiveness of Entropy Weight Method in Decision-Making. Mar 26, 2020, from https:// www. hindawi.com/journals/mpe/2020/3564835/.