

Quantifying the performance of players in a football match

Abhinav Yashkar^{#1}

[#]Student, Department of Civil Engineering, IIT Kharagpur
West Bengal, India

Abstract — *It's hard to evaluate the impact of a player's performance on the team. The tradition method and rating systems involve looking at a few metrics which include goal scored, assists, key passes, tackles, intercept, etc. Often this methodology makes the goal scorer and the assist provider the most important players of the team, which might not always be the case. No wonder Ballon d'Or winners are forwards and not defenders. These numbers make sense when comparing similar metrics. However, when comparing a forward, whose primary job is to score and assist goals, with a defender, whose primary task is to clear the ball, tackle, it's difficult. Football is a team sport; there is a complex interaction between the players. A winning goal might be a result of a threading pass by the midfielder and the winger making a diagonal run to take one of the central defenders with him creating space for the striker to score the goal. In such complicated scenario, it is tough to allocate the contribution of the goal. The paper discussed a statistical method, using regression and optimization, to qualitatively allocate the points contributed to the team by a particular player during a season. Thus, even thou the player who scored the winning goal might have secured 3 points for his team; his contribution to the team is not 3 points. The paper provides a methodology for distributing those 3 points to their rightful contributors.*

Keywords — *Football, Player Performance, Linear Programming Regression, Optimization, player rating*

I. INTRODUCTION

Rating points have replaced the traditional way of gauging a player's impact on the match. Initially, ex football maestros would give their verdict as a part of their football punditry on the post-match analysis shows. But soon technology caught up, and now we have real-time ratings [1]. These ratings are based on various attributes of the players; holistically weighted to come up with a realistic representation of the player's performance [2]. These ratings are excellent at evaluating the player's performance and also comparing players playing in similar positions. However, the impact of the player on the team is not explained by the ratings. However, it is a good starting point. Opta started in 1996 by a group of

friends who used to collect football data watching the game from the pub. Now, Opta is a global company, and it collects data for over 30 sports across 70 countries [3]. Data from opta is collected from websites like Squawka.com and whoscored.com who aggregate the data points to come up with real-time rating point systems.

Rating is tough to interpret because these numbers depend on various factors which are nearly impossible to weigh. For instance, a striker rating is depended on the number of goals he scored, assists, key passes or dribbles. Strikers scarcely get any rating points for tackling, intercepts or ball distribution. A goal or an assist is weighted higher than any other metric. More often than not the striker/attacking midfielder (No. 10) of a team is the highest rated player. Defenders, in general, find it very hard to have high ratings as a clearance or aerial duel does not fetch them as many points as an assist or a goal. Thus, looking at the rating chart for a team will give a false impression of who the most impactful player on the team is [4].

Another flaw of the existing rating system is that the rating system employs a static weighing system. Goal scored is always weighted more than an assist or tackle, or a header won [5]. Also, all the goals are equally weighted. This model is way too simplistic for the game as complicated as football. Consider a team like Barcelona, who regularly score over 2 goals per match (This season, 2016-2017, they are scoring an average of 2.93 goals per match [6]). According to the current rating system if the team scores 3 goals all the goals will be equally weighted. This however, is not the case in a football match. Now consider the example of Villarreal, who are currently in the 5th position of the Spanish League, a great position to be in. They are scoring an average of 1.46 goals per match (as per the current season, 2016-2017). The goal scored by a player of Villarreal is more important than that of Barcelona. Thus, in hindsight, we can say that not all goals, assist, tackles, etc. are equally important [7].

To get an insight into who is the most impactful player on a team, we have to figure out a way to distribute the points won by a team with the players over a period. Points earned by a player is a comparable metric be it a forward or a defender. A novice approach would be to distribute the points weighted by the rating i.e. when a team wins 3

points the player who has a 9/10 rating contributes more to the team than the one who had a rating of 7/10. But this approach is rebuked because of the reason stated above (a striker can easily win more points compared to any other player on the pitch). However, we can use statistical techniques to find the weights to distribution the points among the players. We can also figure out which players perform well under the influence of other players and distribute their share of points to the relevant owners [8]. This method will give us the player who have helped the team accumulate the maximum amount of points and thus, is the most important player for the team.

II. METHODOLOGY

This paper uses linear regression followed by optimization of the estimates such that the estimates obtained from the regression are all positive. Linear Regression is a method for fitting a curve to a set of points using a goodness of fit criterion. The variable we are predicting is called the criterion variable and is referred to as y . The variable we are basing our predictions on are called the predictor variables and are referred to as x_i .

For evaluating the goodness of the fit we use a standard procedure, that is, the least Mean Square (LMS). LMS is a mathematical method for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets (the residuals) of the points from the curve.

The Linear Programming Regression starts with an unconstrained regression. The coefficient from the unconstrained regression may make perfect statistical sense but may not make sense in footballing terms. These abrasions may occur due to limitations of the model, mainly multicollinearity. As mentioned before football is a game of complicated interaction between players. Thus, we need to convert the unconstrained model to a constrained model. In a Linear Programming Regression, constrained model, we put constraints on the coefficients while minimizing the standard error. In this analysis, the cumulative points for a team will be used as the dependent variable, y . The independent variables will be the cumulative rating points, x_i . The model structure will look like as shown in the equation below.

$$\text{Cumulative Score}(y) = a_0 + a_1 * (\text{Player 1 rating score}) + a_2 * (\text{Player 2 rating score}) + \dots + a_n * (\text{Player n rating score})$$

Fig. 1 Model Structure

With the help of the regression, we will estimate the value of a_0, a_1, \dots, a_n . This will be only half the work done. We might get a few estimates as negative. These negative estimates make perfect statistical sense. However, this does not make sense in football i.e. players contribution to the team regarding points is negative. The players whose estimate comes out to be negative can be thought of

as team players. These players do better when the team is performing better, i.e. their performance is based on how the other players on the pitch are performing. It is here where we introduce the optimization of the estimates.

The optimization objective is to minimize the standard error i.e. $\min(\text{standard error} = \sum((\text{actual} - \text{predicted})^2)$. This objective function is subjected to the flowing constrains $l_i < a_i < m_i$, i.e. keeping all the estimates positive. In this equation a_i if the estimate while l_i and m_i are the lower limits and the upper limit. This exercise ensures that all the players have a positive estimate. Finally, we calculate the total points each of the players has contributed to the total point tally. Thus, also confirming the most important players and the players who give a false impression of being valuable to the team; but only play good in support of other players, thus are not impact players.

III. DATA

For this study, we are considering two teams in Manchester United, who play in the English Premier League and Real Madrid who play in the Spanish La Liga. This analysis is for the season campaign from 2014 to 2015. The cumulative team points for Manchester United have been taken from the official website of the English Premier League [9]. The cumulative points for Real Madrid have been taken from the official website of the Spanish La Liga [10]. Individual player ratings have been taken from football statistic and analysis website Squawka [11].

IV. RESULTS

We used two teams for the analysis, Manchester United and Real Madrid season ending 2015.

A. Manchester United

Manchester United, 20 times English Champions, endured a tough season. After the debacle under the former Manager David Moyes last season when the team finished 7th, 4th looked a decent beginning to the era under Louis Van Gaal. Without any analysis, those who watched Manchester United play last year know that there were two outstanding players in David De Gea and Michale Carrick. David De Gea was adjusted as the Player of the season both by the players and the fans for his outstanding goalkeeping. Standing behind a young and nervous defense, David De Gea made phenomenal saves and was almost impermeable. Michael Carrick has been the stalwart of Man United's Midfield for some years now. But his performance this year was exceptional, to say the least. It was not the goals or the assists, but the calmness he brought to an unsettled and nervous team that made him an impactful player. Man United won an impressive 72% of their games with Michale Carrick in the team and won only 35% without him. Since Michael Carrick doesn't score a lot of goals or assists, it is tough to quantify how

much points he won for United. Let us have a look at the results.

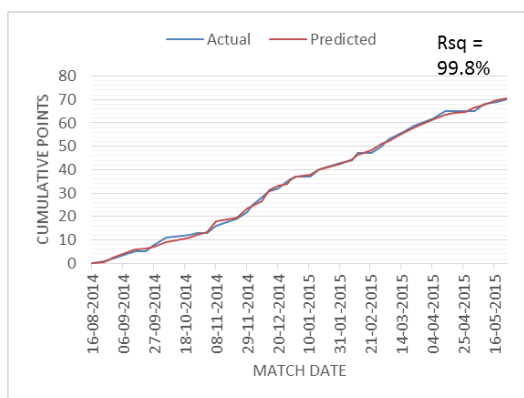


Fig. 2 Manchester United Actual vs. Predicted Points

Fig 1. Shows the actual vs. predicted for the total points accumulated by Manchester United during their campaign of 2014-2015. Tables 1. Shows the point contributed by each of the players.

TABLE I
TOP 10 PERFORMERS FOR MANCHESTER UNITED

Player	Points	Percentage
Michael Carrick	13.3	19.04%
Angel di Maria	10.7	15.27%
Daley Blind	10.3	14.74%
David de Gea	9.6	13.76%
Robin Van Persie	8.2	11.69%
Ander Herrera	7.4	10.52%
Phil Jones	3.9	5.62%
Juan Mata	3.4	4.91%
Wayne Rooney	1.7	2.40%
Marcos Rojo	1.4	2.04%

The names are not at all surprising, as discussed before Michael Carrick was the most influential player. He won 4 matched and a draw for the team that amounts to 13.3 points or almost 1/5th of the points.

Next name on the list is interesting. For many Angel Di Maria had a disappointing season, a season marked by injury, burglary and horrifying performances in the 2nd half of the season. But Angel Di Maria is an impact player. He might not have lived up to the hype of the British record transfer fee, but he was very crucial in the campaign. He had ten assists during his 1645 mins on the pitch. Compare that with Cesc Fabregas, who has the most assists in the Premier League, the assists per mins are almost similar. Cesc Fabregas had an assist every 160 mins, while Di Maria had assists every 164 mins. The numbers show that Di Maria is an impact player and a very crucial part of the team.

The third player on the list is Daley Blind; the former Ajax man is probably the most consistent player during the last campaign. He played a crucial role controlling the game from the back and chipping in with crucial goals. A very versatile player, mostly often overshadowed by the players ahead of him, Daley Blind is a very vital part of the team.

The 4th on the list is David De Gea. Probably one of the best goalkeeper at the moment. David De Gea saved United on numerous occasions. His contribution to the team was three matches won, which is high for any goalkeeper.

B. Real Madrid

When we talk about the Spanish giants, Real Madrid, we talk about big name players, a lot of impact players. Real Madrid of the last season certainly didn't lack the Galactico. The current Ballon d'Or winner Cristiano Ronaldo being the Ace, with the world most expensive player, Gareth Bale and the last world cup golden ball winner James Rodriguez, Madrid has no dearth of big names or quality. Fig 2 shows the Actual vs. Predicted chart for the last campaign. While Table2 shows the point contribution of the players.

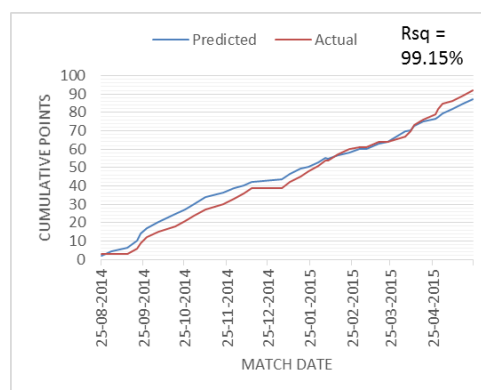


Fig. 3 Real Madrid Actual vs. Predicted Points

TABLE III
TOP 10 PERFORMERS FOR REAL MADRID

Player	Points	Percentage
Cristiano Ronaldo	25.1	27.2%
James Rodriguez	16.7	18.1%
Gareth Bale	14.4	15.6%
Luka Modric	12.3	13.3%
Toni Kroos	11.0	12.0%
Isco	5.8	6.3%
Karim Benzema	4.3	4.7%
Raphael Varane	1.6	1.7%
Marcelo	0.9	0.9%

Cristiano Rolando is the most important player for Real Madrid. He scored an astronomical 48 goals this season which is second to none in Europe. He contributed a total of 25 points to the team last

campaign, i.e., winning eight games. Cristiano Ronaldo himself won almost $\frac{1}{4}$ of the games. There are very few players who can influence the outcome of the game as much as Ronaldo.

The second on the list is the current golden ball winner, James Rodriguez. The Colombian adapted to the team as if he belonged at Madrid. He scored a total of 13 goals and assisted the same number. Not a bad number for a new signing. He contributed a total of 5.5 wins during the last campaign.

The third on the list is Gareth Bale. The most expensive player on the planet had a decent season, but the burden of being the most expensive player makes his number look a bit bleak. Gareth Bale scored some crucial goals, with a tally of total 13 goals and nine assists, he was a crucial part of the team. He contributed five win worth of points last campaign.

In a team like Madrid, who dominate almost any opposition, it is easy to see why the most impactful players are the players who stand up front facing the opposition goal. However, the players who follow the Cristiano Ronaldo, Gareth Bale and James Rodriguez are Luka Modric, Toni Kroos and Isco who are no less valuable to the team as the front three.

V. CONCLUSIONS

The study showed a better way of accessing and comparing player's performance. Rather than analyzing tons of metrics such as goals, assists, key passes, tackles, intercepts, etc. it's better to convert these parameters to contributed points and identify the impactful players. This method also makes sure the points are distributed to the rightful owner. This method is superior to the existing rating methods employed as it facilitates comparing players across positions and teams. This approach also debunks the myth that the goals scorers are the most impactful player on a team. For Manchester United, Michael Carrick and Daley Blind were suggested as the most and the 3rd most impactful players. The goal scorers in Angel Di Maria and Robin Van Persie were ranked 2nd and 5th most impactful.

This approach can also be used to decide which players should be transferred and which not. As in the case of Manchester United, David De Gea contributed almost 9 points, which suggests that the defense has been destitute and needed replacement. Also, it outlines the importance of Michael Carrick and that his replacement has to be a truly world class player. Also, the strikers of Man United haven't been contributing a lot of points as compared to Real Madrid, suggesting the need for a prolific goal scorer. Thus, this study provided a better understanding of the performance of the team than any other individual or a composite metric.

REFERENCES

- [1] R. D. Baker and I. G. McHale, "Time Varying rating in association football," *Journal of the Royal Statistical Society*, pp. 481–492, 2015.
- [2] I.G. McHale, P.A. Scarf and D.E. Folker, "On the development of a soccer player performance rating system for the English Premier League" *Interfaces*, vol. 42, No. 4, pp. 339-351, July 2012.
- [3] Theguardian article on Opta. [Online]. Available: <https://www.theguardian.com/football/blog/2011/aug/10/opta-stats-premier-league>
- [4] J. Albert and J. Bennett, *Curve Ball*, New York: Springer, 2003.
- [5] D. Hiscock, B. Dawson, J. Heasman, and P. Peeling, "Game movement and players performance in the Australian Football League," *International Journal of Performance Analysis in Sport*, vol. 12, pp. 531–545, 2012.
- [6] Whoscored homepage. [Online]. Available: <https://www.whoscored.com/>
- [7] C. Anderson and D. Sally, *The number game: Why everything you know about football is wrong*, Penguin, 2013.
- [8] I. McHale and P. Scarf, "Modelling soccer matches using bivariate discrete distribution with general dependence structure" working paper series, Salford Business School, Manchester, United Kingdom, Feb. 2007.
- [9] English Premier League homepage. [Online]. Available: <https://www.premierleague.com/>
- [10] Spanish LaLiga homepage. [Online]. Available: <http://www.laliga.es/en>
- [11] Squawka homepage. [Online]. Available: <http://www.squawka.com/home/>