Honest Mirror: Quantitative Assessment of Player Performances in an ODI Cricket Match

Madan Gopal Jhawar[1]⋆ and Vikram Pudi[2]

[1] Microsoft, India
majhawar@microsoft.com
[2] IIIT Hyderabad, India
vikram@iiit.ac.in

Abstract. Cricket is one of the most popular team sports in the world. Players have multiple roles in a game of cricket, predominantly as batsmen and bowlers. Over the generations, statistics such as batting and bowling averages, and strike and economy rates have been used to judge the performance of individual players. These measures, however, do not take into consideration the context of the game in which a player performed. Furthermore, these types of statistics are incapable of comparing the performance of players across different roles. In this paper, we present an approach to quantitatively assess the performances of individual players in single match of One Day International (ODI) cricket. For this, we have developed a new measure, called the Work Index, which represents the amount of work that is yet to be done by a team to achieve its target. Our approach incorporates game situations and the team strengths to measure the player contributions. This not only helps us in evaluating the individual performances, but also enables us to compare players within and across various roles on a common scale. Using the player performances in a match, we predict the player of the match award for the ODI matches played between 2006 and 2016. We have achieved an accuracy of 86.80% for the top-3 positions, which is superior to baseline models and previous works, to the best of our knowledge.

Keywords: Cricket Analytics, Player Performances, Player of the Match

1 Introduction

Cricket is majorly played in three formats – Test, ODI and Twenty20 (T20), with ODI being one of the most followed formats. An ODI is a form of limited overs cricket, played between two teams where each team has a combination of batsmen and bowlers making up to 11 players in total. Each team bats for a maximum of 50 overs where an over is defined as a set of six deliveries bowled by the bowlers of the opponent team. An ODI cricket match starts with a coin toss and the Captain of the side winning the toss chooses to either bat or bowl first. The team batting first sets the target score in a single innings, where the

⋆ This work was done when the author was a student at IIIT-Hyderabad.

innings lasts until the batting side loses all the 10 wickets, where a wicket refers to a player getting out, or the batting side's quota of 50 overs is completed. The team batting second tries to score more runs than the target score in order to win the match. Similarly, the side bowling second tries to take all the 10 wickets of the opponent team or make them exhaust their overs before they reach the target score in order to win.

In a game of cricket, the batsmen of one team play against the bowlers of the other team, and vice-versa, in order to win the match. Therefore, evaluating the performances of individual players in a game of cricket becomes very critical. It helps in segregating the players who are contributing to the team from the ones who are failing to deliver on the ground. However, evaluating the performances of players is not a straight-forward task. Traditionally, statistics such as batting and bowling averages, and strike and economy rates have been used to assess the performance of individual players. However, these statistics fail to incorporate several important aspects of the game. Runs scored or wickets taken under pressure at crucial stages are of more value as compared to scoring more number of runs or taking more number of wickets. Furthermore, assessing the overall performance of an individual cricketer requires a comprehensive evaluation of his contributions to the team, both in terms of his batting and bowling contributions. However, combining and comparing the batting and bowling performances of a player, on a common scale, is a challenging task and often becomes a subjective decision.

Therefore, in this paper, we propose a methodology to quantitatively assess the performances of individual players in a single game of ODI cricket match. We introduce a new measure, called the Work Index, which represents the amount of work yet to be done by a team to reach their expected target score. Work Index incorporates several important aspects of a game, including the current stage of the match, the progress so far as compared to the initial estimations, the two competing teams' strengths, etc. We measure the Work Index for both the batting as well as bowling teams, namely, the Batting Work Index and the Bowling Work Index. The former denotes the amount of work to be done by the batting team to reach the target, while the latter represents the amount of work to be done by the bowling team to restrict the batting team from reaching the target. Using these two work indices, we calculate a contribution-score for each player which incorporates his batting and bowling contributions towards achieving the team's overall goal.

Furthermore, akin to many other sports, the player of the match title is awarded to the player who played the most significant role in a match of cricket. Today, it is chosen by the match committee and the commentators which makes it a subjective decision. Therefore, we propose a methodology to determine the player of the match using the player contribution-scores calculated by our approach. We compare our model with previous works and other baseline models, and the superiority of our model over others further proves the validity of our approach.

2 Related Work

In literature, Duckworth and Lewis proposed a real-time measure to estimate the amount of resources remaining with a team, called as D/L resources, as a function of number of balls and wickets remaining. They further used the D/L resources to reset targets in rain interrupted matches [1]. It is said to be one of the most pioneering works in cricket analytics and was adopted by the International Cricket Council (ICC) in 1998.

Pertaining to assessing player performance, Johnston et al. [4] used dynamic programming formulation to develop a method of calculating the contribution, in runs, made by each player to the team's score in a game of one-day cricket. Lewis [2] used Duckworth/Lewis methodology to create alternative measures of player performances in a game of cricket. These measures take into account the stages of innings when runs are scored or conceded and wickets are taken or lost.

Recently, Bhattacharjee et al. [5] proposed a measure of quantifying the pressure, named as *Pressure Index*, on the teams batting or bowling in limited overs cricket matches. They use D/L resources, as proposed in [1], ratio of the wickets lost and the current as well as the initial required run rates to quantify the pressure on a team. Further, they use the pressure index to access the individual player performances in a specific match. However, the method could be used to quantify the pressure on a team only for the second innings of a match, where the batting team has a fixed target to chase. Also, their approach takes into account the ratio of the wickets fell down instead of incorporating the varying strengths of individual players. This is a very critical factor because teams do not play with a fixed number of specialized batsmen. Losing 6 wickets has a different impact on a team playing with 6 specialized batsmen as compared to a team playing with 7 specialized batsmen. Furthermore, no quantitative method, in any form, of validating the approach has been discussed.

Therefore, in this paper, we propose a new dynamic measure, called the *Work Index*. Apart from the runs scored and balls bowled, the work index also incorporates the potentials of the batsmen and bowlers who are remaining to perform. In addition to this, with the use of D/L resources, we propose a method which enables us to calculate the work index in the first innings also, where a team does not have a fixed target to chase. We further use the work index to quantitatively assess the player performances in a match and predict the player of the match.

3 Methodology

Our methodology to assess the player performances for a given ODI cricket match involves estimating a new measure, called the *Work Index*. Work Index incorporates several crucial qualitative and quantitative aspects of the game. The **three parameters** considered in calculating the work index are as follows:

- The progress, in terms of runs scored, towards chasing the set target.

- The current stand, in terms of scoring rate, of the batting team relative to the initial estimations.
- The remaining batting and bowling potentials of the batting and bowling teams, respectively.

The first and second parameters capture the quantitative aspects of the game situation in terms of the runs scored and the required run rate as compared to the initial required run rate for the batting team. On the other hand, the third measure captures the quality of the batsmen and bowlers remaining for the batting and bowling teams, respectively. Therefore, work index, a blend of these features, captures the context at a given stage of the match.

3.1 Target Estimation in First Innings

Calculating the first and second parameters requires us to know the target score the batting team is trying to achieve. In an ODI cricket match, the team batting second has a predefined target, set by the opponent team, to chase in order to win the match. On the other hand, the team batting first does not have a fixed target to score. Ideally, the team would try to score as many runs as possible, but there are limited resources in terms of the number of overs and wickets. Therefore, at the start of the first innings, we estimate the target score to be the average number of runs scored in the first innings of all the matches played in the same country in the past, where the team batting first was able to successfully defend their score. This estimated target score is further improved after every ball is bowled as per Equation 1, depending upon the actual situation of the match.

$$newTarget \leftarrow runsScored + DLrem * initTarget \tag{1}$$

where $runsScored$ is the number of runs scored by the batting team at the current stage of the game, $DLrem$ is the ratio of the D/L resources remaining with the team [1][7], and $initTarget$ is the target estimated at the start of the innings.

With a defined target score to achieve for both the innings, Equations 2 and 3 represent the mathematical formulation of the first and second parameters of Work Index (as mentioned at the start of Section 3), respectively.

$$k \leftarrow runsRemaining/Target \tag{2}$$

$$r \leftarrow reqRunRate/initRunRate \tag{3}$$

where $runsRemaining$, $reqRunRate$ and $initRunRate$ represent the number of runs yet to be scored, the average number of runs required per over from the current stage and the average number of runs require per over at the start of the match, respectively, by the batting team to achieve its target score. The higher values of k and r intuitively tell us that the batting team has a lot of work to do to reach the target score, and vice-versa.

3.2 Modeling Players and Teams

Calculating the third parameter requires us to model the player and team potentials for a given match. Therefore, as proposed in [3], we use the *Batting Strike Rate* and *Batting Average* to calculate a player's batting score, and *Bowling Strike Rate* and *Bowl Economy* to calculate his bowling score, where *Batting Strike Rate* is the average number of runs scored per 100 balls faced, *Batting Average* is the average number of runs scored per innings, *Bowling Strike Rate* is the average number of balls bowled per wicket taken, and *Bowling Economy* is the average number of runs conceded per over by the player.

Further, we define the *total batting score* of a team to be the summation of the batting scores of all of it's players. And similarly for the *total bowling score* of a team. However, at a given state of the match, some of the players from the batting team would have got out and some of the players from the bowling team would have bowled a part of their quota of maximum 10 overs (60 balls). Therefore, we define the *remaining batting score* of the batting team as the sum of the batting scores of only those players who haven't got out yet. Similarly, the *remaining bowling score* of the bowling team is calculated as the sum of all the individual bowlers' bowling scores, weighed by the ratio of number of balls he has remaining to bowl in this match to the maximum number of balls he can bowl in an ODI cricket match, i.e., 60.

Hence, with a defined method to estimate the total and remaining batting and bowling potentials of the batting and bowling teams, respectively, Equations 4 and 5 represent the mathematical formulation of the third parameter.

$$b \leftarrow remBatScore/totalBatScore \tag{4}$$

$$w \leftarrow remBowlScore/totalBowlScore \tag{5}$$

Where $totalBatScore$ and $remBatScore$ represent the total and remaining batting potentials of the batting team. And similarly for the $totalBowlScore$ and $remBowlScore$. The higher values of b and w tell us that the batting and bowling teams have got good players to perform who can win the match for them, respectively.

3.3 Work Index

Having formalized all the three parameters of work index, we will now explain the Work Index in detail. Work Index is a dynamic measure which is updated as the innings progresses and takes into account the current state of the match to estimate the work yet to be done. With the help of variables defined in Equations 2, 3, 4 and 5, the batting and bowling work indices are calculated as per the Equations 6 and 7, respectively.

$$\mathbf{battingWorkIndex} \leftarrow 100 * k * (r + \alpha * (1 - b) + \beta * w) \tag{6}$$

$$\mathbf{bowlingWorkIndex} \leftarrow 100 * k * (1/r + \alpha * b + \beta * (1 - w)) \tag{7}$$

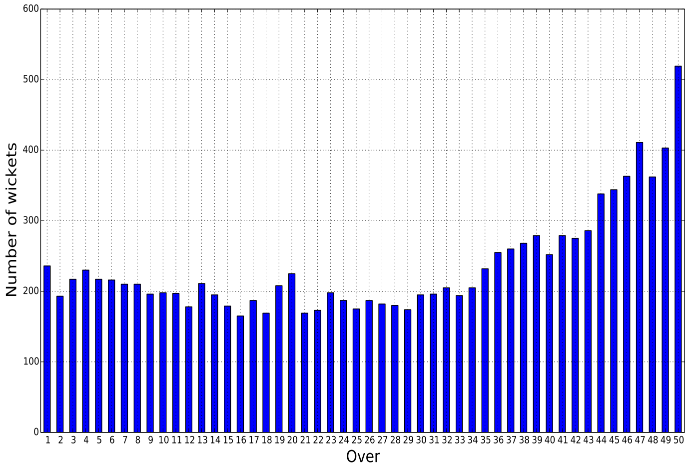

Fig. 1. Total number of wickets fell in each over in our dataset. As a game reaches its final stages, the batsmen adopt a riskier strategy to score runs at a higher rate, with less concern about losing wickets.

Variable k (defined in Equation 2), defined as the ratio of the runs remaining to score with respect to the target, is used as a bias in calculating the work index. The lower values of k, generally found at the end of the innings, directly reduces the impact of the other factors in determining the work index. As it can be seen from Figure 1, as a game reaches its final stages, the batsmen adopt a riskier strategy to score runs at a higher rate, with less concern about losing wickets. The value of a wicket reduces as scoring runs becomes the sole purpose. Similarly, higher values of k, found at the start of the innings, boosts the impact of other factors. This is because, at the initial stages of the innings, the wickets of the batsmen carry a lot more importance. Losing early wickets at the start of an innings puts the batting team into tremendous pressure, as they lose their key batsmen and face the threat of getting all-out, before even completing the quota of 50 overs. Variable r (Equation 3) captures how well is the batting team scoring as compared to the initial estimations. It is directly proportional to the batting work index, as increased required run rate, increases the amount of work to be done, and similarly it is inversely propotional to the bowling work index.

Variables b and w (Equations 4 and 5) represent for the remaining batting and bowling potentials of the corresponding teams. They account for the amount of batting and bowling resources remaining with the batting and bowling teams, respectively. Incorporating individual player's skills into these variables enables us to assess the current game scenario in a detailed way. They enable us to capture those scenarios where a team has lost several wickets yet still has good players remaining in the batting line-up, who can potentially change the game's direction. The parameters, α and β, represent the relative weightage of the remaining batting potential and the remaining bowling potential, respectively. Also, bowlers have 300 balls to possibly get the batsmen out, whereas the batting team possesses only 10 wickets to score runs. Therefore, losing wickets has significant impacts on the batting team. The values of the parameters, α and β, have been discussed in the experiments Section 4.

In all, work index is a combination of many important aspects of the game that enables it to capture the overall game scenario.

3.4 Player of the Match

Cricket is a game of bat and ball. Bowlers from the bowling team take turns in overs to bowl their quota of 50 overs, where, in an over, one of the players from the bowling team bowls 6 deliveries to the batsman at strike. We calculate the *Batting Work Index* and the *Bowling Work Index* after each ball is bowled. The difference between the two consecutive batting work indices is contributed to the batsman who played the corresponding ball. Similarly, the difference between the two consecutive bowling work indices is contributed to the bowler who bowled the ball. We repeat this process for each ball bowled in the entire match. At the end of the match, the scores attributed to an individual player represents his all-round performance in the entire match. This score not only captures the quantitative amount of runs scored or wickets taken by a player, but also the context in which he made the contributions. Phrases like *"Catches win Matches"* are very popular in the game of cricket and therefore, the fielding efforts of players could also be integrated to capture the players' overall contribution to the match. However, we have tabled it for the future work as of now.

Furthermore, *Player of the Match* title is awarded to the player who played the most significant role in a particular match. At first, it seems that the player who has the maximum score at the end of the match should be awarded the player of the match. However, player of the match award is almost always (95.83% of the times) given to a player from the winning side. A player from the losing side bags the player of the match award only if his performance is significantly better than the others and has contributed to an almost win for the losing side. Therefore, as of now, we choose the players from the winning side only as the potential candidates for the player of match award. We rank them directly based on their scores and the player with the maximum score is awarded the player of the match. Note that some heuristics can be used to capture the very rare player of the match awards from the losing side also. However, for now, we have tabled it for future work.

4 Experiments and Results

We have studied all the ODI cricket matches played between 1st of January, 2006 and 30th June, 2016. Ball-by-ball data for each match has been taken from the cricsheet database [8]. We have focused our study to only the top 9 ODI-playing teams, namely, India, Australia, South Africa, England, Sri Lanka, Pakistan, New Zealand, Bangladesh and West Indies. Since the impact of nature on the game cannot be foreseen, a total of 216 matches which were either interrupted by rain or ended up in a draw/tie, have been removed from the dataset. Finally, we studied a total of 786 ODI cricket matches.

For the potential player of the match candidates, as discussed in Section 3.4, we consider only the players from the winning team. For a given match of cricket, our model outputs a list of player ranked in the descending order of their contribution in the match. Hence, to measure the efficiency of our model, we use an exponential-ranking metric. For a given match, the match score is calculated to be $1/2^{R-1}$, where $1 \leq R \leq 11$ is the rank at which the player of the match has been predicted by our model. Therefore, the exponential-rank of our model is calculated as the summation of the match scores for all matches. The choice of an exponentially decaying metric over existing metrics such as *Mean Reciprocal Rank* has been made to increase the penalty for wrong predictions, as there is only one player of the match as compared to top-k relevant outputs.

With the defined accuracy metric, we use a validation set containing all the matches played between January, 2006 and December, 2012 to find the most suitable values of the parameters α and β, defined in Equations 4 and 5. Table 1 tabulates the exponential-rank of our model on the validation set for multiple values of the parameters. As it can be seen, $\alpha = 2.0$ and $\beta = 1.5$ yields the best results. Therefore, these values will be considered for the further discussions.

Table 1. Exponential rank of *Honest Mirror* for multiple values of the parameters, α and β, on the validation set. $\alpha = 2.0$ and $\beta = 1.5$ yields the best results.

$\alpha\beta$	*0.5*	*1.0*	*1.5*	*2.0*	*2.5*
0.5	362.1	357.9	346.2	334.3	330.1
1.0	369.9	371.2	367.8	354.7	344.2
1.5	349.8	379.6	375.9	373.1	360.8
2.0	328.7	363.7	**384.8**	378.7	371.9
2.5	306.1	344.7	372.6	379.8	377.5

Accuracy (in %) for player of the match for the first 5 ranks are shown in Figure 2. A decreasing curve proves that the players who are performing better are placed higher in the rankings than the others. We have achieved 59.14% accuracy for the first rank and an accuracy of 77.79% and 86.80% for the top two and top three ranks respectively.

In literature, to the best of our knowledge, we could not find any previous work on predicting the player of the match for ODI cricket matches. However, Bhattacharjee et al. [5] proposed a model, the *PI Model*, to assess player performances in a game of limited overs cricket match using pressure index, but only for the second innings of a match. We extended their method for the first innings by estimating the target score using the same approach as discussed in Section 3, and add up the player's batting and bowling contributions for both the innings to calculate a player's overall performance in a match. The players from the winning team are considered to be the potential player of the match in the order of their overall contribution in the match. We implemented their work, to the best of our abilities, to compare their approach against our model.

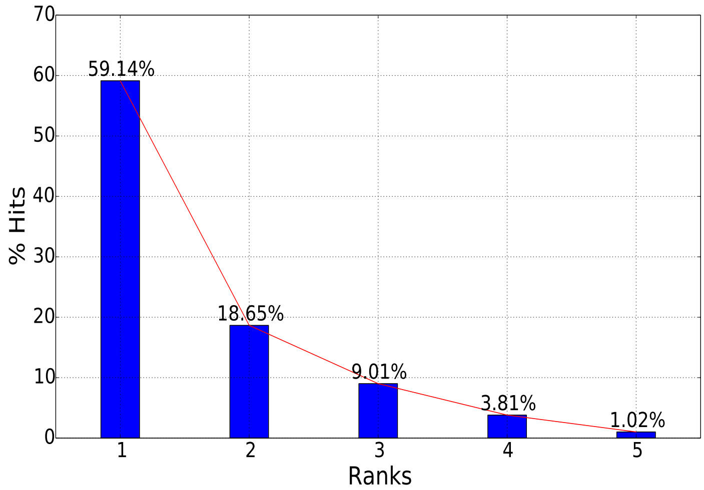

Fig. 2. Accuracy (in %) for player of the match for the first 5 ranks. A decreasing curve proves that the players who are performing better are placed higher in the rankings than the others.

Apart from that, in a game of cricket, the number of runs scored by a player and the number of wickets taken by him are the two major criterion to judge a player's performance. Therefore, we further compared our approach with the two following baseline models which take into account a player's overall contribution in a match–

– **Model_1**: The overall contribution of a player is the summation of the ratio of the runs he has contributed to the teams total batting score and the ratio of the wickets he has taken to the total of wickets taken by his team.
– **Model_2**: To be able to combine the runs scored and wickets taken by a player, we map one of these into another, i.e., we calculate the weight of a wicket taken by a bowler in terms of the runs scored by a batsman. The weight of a wicket, denoted by ω, is calculated as the total number of runs scored in the match divided by the total number of wickets fell down. Therefore, the total contribution of a player in a match is the summation of the number of runs scored by him and ω times the number of wickets taken by him.

Figure 3 demonstrates the exponential-rank comparison for the top-5 positions for the four models. The number of right predictions, i.e., at the first position, is higher by our model as compared to the others. Hence, the superiority of our model against the others validates our approach.

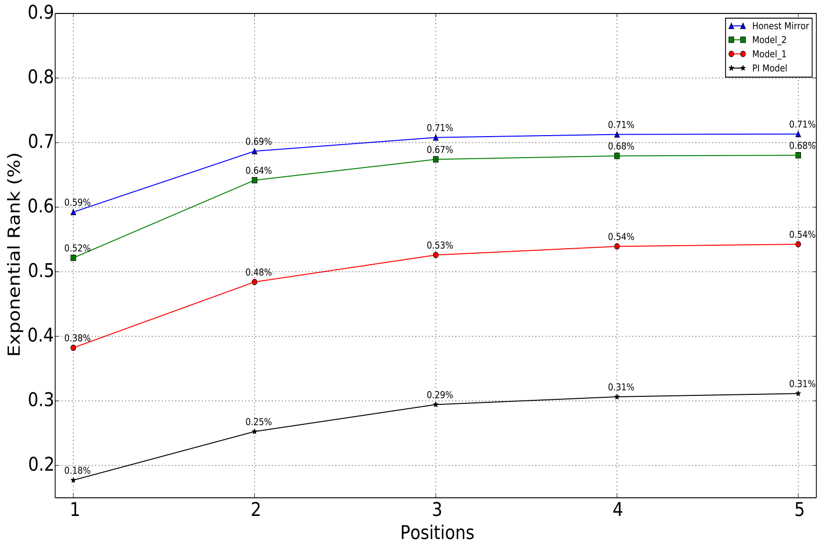

Fig. 3. Exponential-rank for the the top 5 positions for all the four models. Higher value for *Honest Mirror* at the first rank proves that our model is able to pick the player of the match by taking the game situations into account.

References

1. Duckworth, Frank C., and Anthony J. Lewis. "A fair method for resetting the target in interrupted one-day cricket matches." Journal of the Operational Research Society 49.3 (1998): 220-227.
2. Lewis, A. J. "Towards fairer measures of player performance in one-day cricket." Journal of the Operational Research Society 56.7 (2005): 804-815.
3. Barr, G. D. I., and B. S. Kantor. "A criterion for comparing and selecting batsmen in limited overs cricket." Journal of the Operational Research Society 55.12 (2004): 1266-1274.
4. Johnston, Mark I., Stephen R. Clarke, and David H. Noble. "Assessing player performance in one-day cricket using dynamic programming." (1993).
5. Bhattacharjee, Dibyojyoti, and Hermanus H. Lemmer. "Quantifying the pressure on the teams batting or bowling in the second innings of limited overs cricket matches." International journal of Sports Science Coaching 11.5 (2016): 683-692.

6. Bailey, Michael, and Stephen R. Clarke. "Predicting the match outcome in one day international cricket matches, while the game is in progress." Journal of Sports Science Medicine 5.4 (2006): 480.
7. DL Table `http://www.tcuandsa.org/Doc/dldocs/DLResourceChartNew.pdf`
8. Cricsheet `http://cricsheet.org/`