Real time measurement of individual influence in T20 cricket

MathSport International 2017

Ankit K. Patel, Paul J. Bracewell & Jason D. Wells

JULY 2017

Presentation Outline

1. Cricket background
2. Research objectives
3. Key findings
4. Data
5. Research challenges
6. Research methodologies
7. Model applications + results
8. Concluding remarks
9. References + Questions

The Game

Overview of Cricket

- Cricket is a bat and ball game played between two teams of eleven on a cricket field (Figure 1).
- Objective: Accumulate more runs than the opposition
  a. For any given innings, the batsmen's objective is to score as many runs as possible given the allocated set of resources (i.e. balls and wickets)
  b. For any given innings, the bowlers objectives are to prevent the scoring of runs and to dismiss the batsmen (i.e. depleting batsmen resources as soon as possible).
- The winning team is the one that scored the most runs

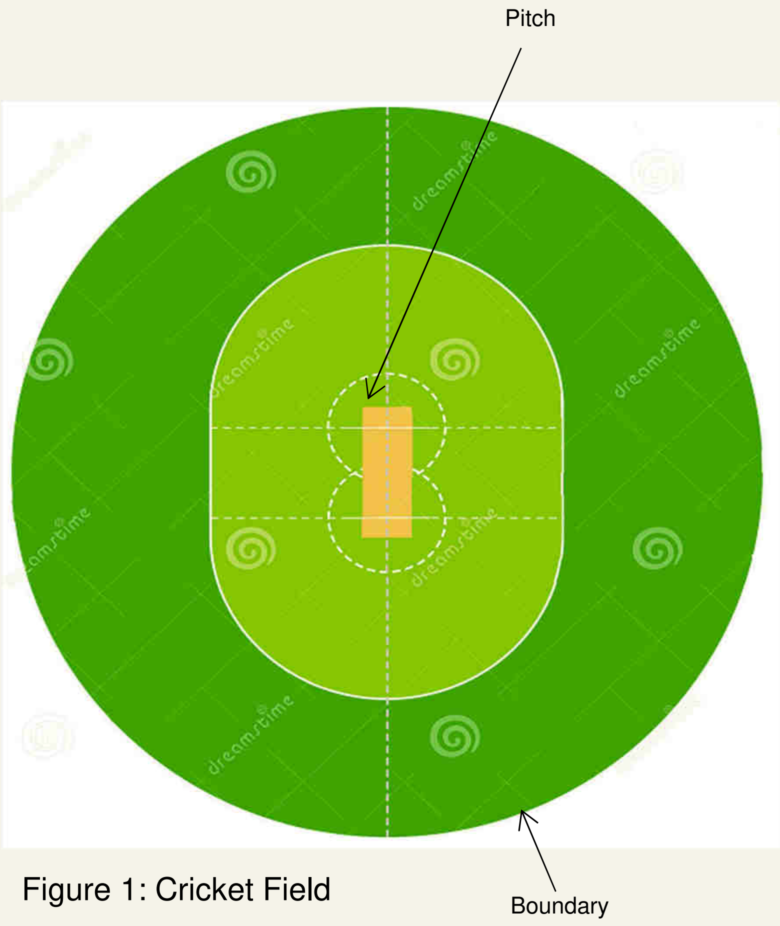


Figure 1: Cricket Field

Research objective

Overall goal

Develop a real-time predictive rating system that measures the amount of influence a player exerts on a Twenty-twenty match of cricket during any stage of an innings.

Key findings

Overall outcome

Real-time application

Model outputs (i.e. ball-by-ball influence scores) were indicative of an individual's ability to influence match outcome.

Static application

Averaging a players ball-by-ball match influence produced values that were indicative of players worth (i.e. player ratings).

Data features

Extraction and volume

R-scripts programmatically extracted ball-by-ball data from ESPNCricinfo commentary logs:

a. 377 Twenty-twenty matches

b. Removed rain interrupted matches

c. ~95,000 observations

d. T20 competitions: Indian Premier League (2014, 2015, 2016), Caribbean Premier League (2014, 2015, 2016), English NatWest (2015, 2016) and Australian Big Bash (2014, 2015)

Research challenges

Definitional and metric challenges

1. Define a player's match influence
    a. Influence = Volume of contribution + Efficiency of contribution + Contribution under pressure
2. Identify volume, efficiency and pressure based metrics
3. Conventional performance metrics are either:
    a. Unobservable on a ball-by-ball basis
    b. Undefined during various stages of an innings

Research methodology

Given the research challenges the following methodology was implemented:

1. Understand the value a player contributes on a ball-by-ball basis.
   - *Calculated the expected runs scored on a ball-by-ball basis*
2. Expected runs dependent on resources (i.e. ball remaining and wickets lost)
   - *T20 Ball-by-ball resources table (Bhattacharya, Gill & Swartz, 2011)*
3. Understand how a batsmen's action affects their probability of dismissal
   - *Batsmen Survival Probability (Brown, Patel & Bracewell, 2016)*

These 3 metrics will enable the creation of volume, efficiency and pressure based metrics on a ball-by-ball basis.

T20 ball-by-ball resource table

Modified Duckworth-Lewis

Bhattacharya, Gill & Swartz (2011)

1. Isotonic regression

$$F = \min_{y_{uw}} \sum_{u=1}^{20} \sum_{w=0}^{9} q_{uw}(r_{uw} - y_{uw})^2 \; s.t.,$$

$$y_{u,w} \geq y_{u,w+1} \; and \; y_{u,w} \geq \; y_{u-1,w}$$
$$y_{20,0} = 100;$$
$$y_{0,w} = 0 \; for \; w = 0, \ldots, 9;$$
$$y_{u,10} = 0 \; for \; u = 1, \ldots, 20;$$

where,

$$r_{u,w} = \frac{mean[x(u,w)]}{mean[x(20,0)]}$$

T20 ball-by-ball resource table cont...

Modified Duckworth-Lewis

2. Bayesian Model – Posterior density

$$exp\left\{-\frac{1}{2}\sum_{u=1}^{20}\sum_{w=0}^{9} q_{uw}(r_{uw}-y_{uw})^2\right\}$$

3. Gibbs Sampling – Full conditional

$$[y_{uw}|.]\sim Normal\left(r_{uw},\frac{1}{q_{uw}}\right)$$

	Wickets Lost						
Overs Available	0	1	2	3	4	5	6
20	100.0	96.8	92.6	86.7	78.8	68.2	54.4
19	96.1	93.3	89.2	83.9	76.7	66.6	53.5
18	92.2	89.6	85.9	81.1	74.2	65.0	52.7
17	88.2	85.7	82.5	77.9	71.7	63.3	51.6
16	84.1	81.8	79.0	74.7	69.1	61.3	50.4
15	79.9	77.9	75.3	71.6	66.4	59.2	49.1
14	75.4	73.7	71.4	68.0	63.4	56.9	47.7
13	71.0	69.4	67.3	64.5	60.4	54.4	46.1
12	66.4	65.0	63.3	60.6	57.1	51.9	44.3
11	61.7	60.4	59.0	56.7	53.7	49.1	42.4
10	56.7	55.8	54.4	52.7	50.0	46.1	40.3

Figure 2: T20 Resource table extract

Expected runs model

Calculate expected total during any stage of a match

1. Gradient Boosted Machine
   a. First innings total = total wickets + total balls + team strike rate + team % boundaries + projected total + current run rate + resources available
   b. Second innings total = team runs + remaining total + total wickets + total balls + resources available + team % boundaries + team % dots + team strike rate + current run rate

Model	Accuracy Measurements			
	Correlation	R^2	RMSE	MAE
Model 1	0.73	0.53	18.6	14.6
Model 2	0.57	0.33	20.1	16.6

Table 1: Expected runs model accuracy

Expected runs model cont…

Understand the value a player contributes on a ball-by-ball basis

2. Features created

$$Batsmen\ runs\ contributed_i = \begin{cases} positive, if\ expected\ runs_i - expected\ runs_{i-1} > 0 \\ negative, if\ expected\ runs_i - expected\ runs_{i-1} \leq 0 \end{cases}$$

$$Bowler\ runs\ saved_i = \begin{cases} positive, if\ expected\ runs_i - expected\ runs_{i-1} \leq 0 \\ negative, if\ expected\ runs_i - expected\ runs_{i-1} > 0 \end{cases}$$

Batsmen survival probability

Understand how a batsmen's action affects their probability of dismissal

Brown, Patel & Bracewell (2016)

1. Investigated the likelihood of a first innings opening batsmen surviving each ball faced
2. Cox proportional hazard models
3. Features created
    a. Ball-by-ball survival probability
    b. Cumulative area under the curve = batsmen current rating

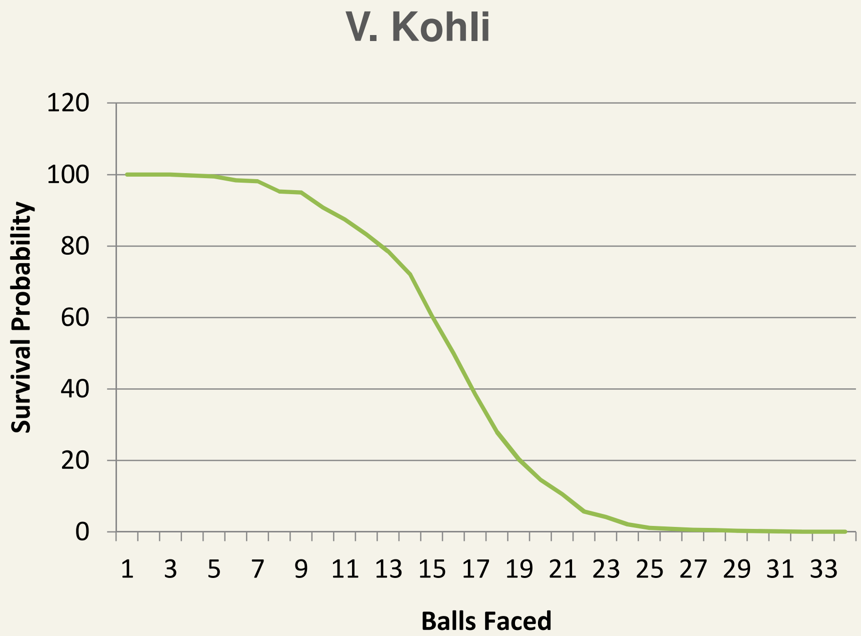


Figure 3: V. Kohli survival curve

Influence model creation

Research challenges + methodology

- ✓ Understand the value a player contributes on a ball-by-ball basis.
- ✓ Expected runs dependent on resources (i.e. ball remaining and wickets lost)
- ✓ Understand how a batsmen's action affects their probability of dismissal
- ✓ Created observable and defined ball-by-ball metrics

Player influence model

Match Outcome

1. Ball-by-ball observation allocated 0 or 1 indicating loss/ win, respectively.
2. Training set broken into four sets:
    a. First innings batting
    b. Second innings batting
    c. First inning bowling
    d. Second inning bowling
3. Applied Naïve Bayes and Logistic Regression models to identify metrics that have a practically and statistically significant affect on match outcome.

Player influence model cont…

Naïve Bayes + Logistic Regression

1. The 4 model incorporate volume, efficiency and pressure based metrics.
2. Modes produce probabilities of winning, given a players current performance.
3. The probabilities are multiplied by 100 to generate a player rating measure.

Role Type	Innings	
	First	**Second**
Batsmen	Area under the curve	Area under the curve
	Runs scored	% Dots
	Total runs contributed	Total runs contributed
	% contribution	Strike rate
	% Dots	%Dots
Bowler	Total runs saved	Total runs saved
	Runs conceded	Runs conceded
	Economy rate	Economy rate
	% boundaries	% Dots

Table 2: Significant variables for influence model

Model application + results

Australian Big Bash League 2016

Real-Time [ball-by-ball] application

i. Binary transform the ball-by-ball influence:

$$Transformed\ Influence_i = \begin{cases} 1, if\ influence\ rating_i \geq 50 \\ 0, if\ influence\ rating_i < 50 \end{cases}$$

ii. Performance measures

- Classification rate = 57%; AUC = 0.60 (overs 1- 20)
- Classification rate = 60%; AUC = 0.65 (overs 5-20)

Model application + results

Australian Big Bash League 2016

<u>Static application</u>

i. Avg. match influence = player rating

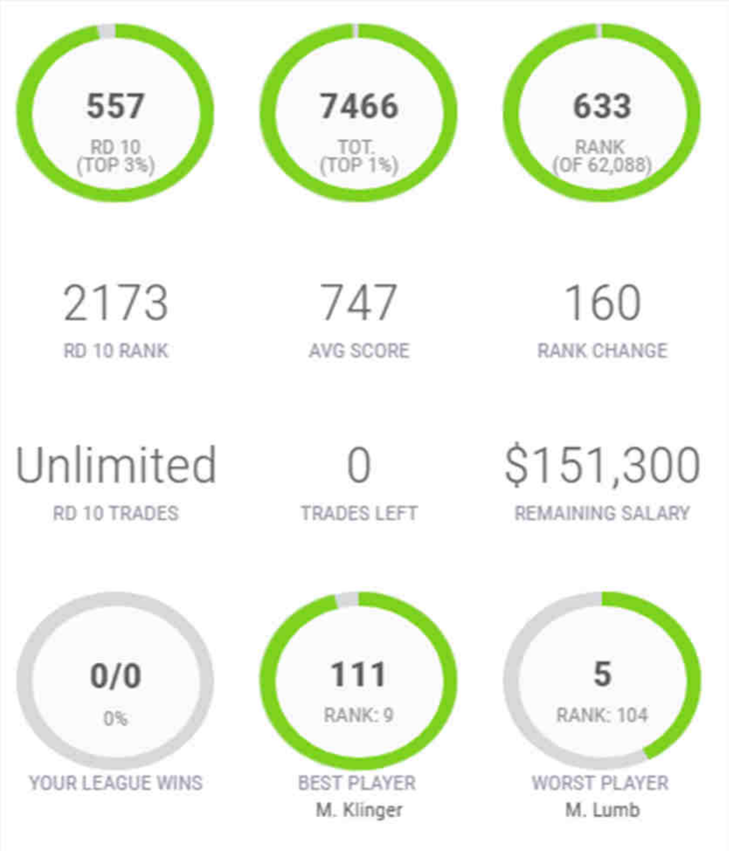
ii. Big Bash Fantasy League

iii. 8 rounds; 35 games

iv. Optimization method to select teams on a round by round basis

v. Finished in the top 1%

vi. 15% increase in team salary



Concluding remark

Key takeaways

1. Identified a novel method to dynamically assess player performance and match influence

2. Create player specific training regiments

3. Model approach can be used to optimize team selection, scout youth talent, develop players and in-game strategies

4. The influence models have recently been updated and built into a DOT product (Mr. Clutch v3) – this paper is product documentation v1.

References

1. Bhattacharya, R., Gill, P. S., & Swartz, T. B. (2011). Duckworth–Lewis and twenty20 cricket. *Journal of the Operational Research Society*, *62*(11), 1951-1957.
2. Brown, P., Patel, A.K., & Bracewell, P.J. (2016, July 12). Real Time Prediction Of Opening Batsmen Dismissal in Limited Overs Cricket. Paper presented at The Proceedings of the 13th Australian Conference on Mathematics and Computers in Sports. (pp. 80-85). Melbourne, Victoria, Australia: ANZIAM MathSport. ISBN: 978-0-646-95741-8
3. Clarke, S. R. (1988). Dynamic programming in one-day cricket-optimal scoring rates. *Journal of the Operational Research Society*, *39*(4), 331-337.
4. Davis, J., Perera, H., & Swartz, T. B. (2015). A simulator for Twenty20 cricket. *Australian & New Zealand Journal of Statistics*, *57*(1), 55-71.
5. Swartz, T. B., Gill, P. S., & Muthukumarana, S. (2009). Modeling and simulation for one-day cricket. *Canadian Journal of Statistics*, *37*(2), 143-160.