

# Stats Aren't Everything

## Learning Strengths and Weaknesses of Cricket Players

Swarup Ranjan Behera and Vijaya V Saradhi

Indian Institute of Technology Guwahati  
Guwahati, Assam, India  
{b.swarup,saradhi}@iitg.ac.in

**Abstract.** Strengths and weaknesses of individual players are understood informally by players themselves, coaches, and team management. However, there is no specific computational method to obtain strengths and weaknesses. The objective of this work is to obtain rules describing the strengths and weaknesses of an individual player. Instead of looking at the traditional statistics, which are nothing but the raw counts of certain events in the game, we focus on cricket text commentaries, which are written narratives giving a detailed description of a minute-by-minute account of the game while it is unfolding.

**Keywords:** Text mining · Sports analytics · Dimensionality reduction.

## 1 Introduction

Cricket is a sport played by tens of thousands and loved by billions of people from all corners of the globe. Cricket is known for recording every detail. A huge amount of data in the form of scorecards, video broadcasts, commentary, and coverage articles is generated in every match. Statistics have been employed widely for years on the generated data. While interesting, the focus of the analysis has primarily been at an aggregate level. Cricket text commentary, on the contrary, is a rich source of fine-grained details about each delivery of the game. Consider an example of short text commentary:

*“106.1, Anderson to Smith, 1 run, 144 kph, England have drawn a false shot from Smith! well done. good length, angling in, straightens away, catches the outside edge but does not carry to Cook at slip.”*

This commentary describes the first delivery in the 107<sup>th</sup> over of the game. Bowler Anderson has bowled this delivery to the batsman Smith. The outcome of the ball is one run. The speed of the delivery is 144 kph (kilometer per hour). The rest of the text is unstructured and describes the way the ball is delivered, and the way batsman played it. For instance, this commentary describes several features of bowling, such as length (good length), and movement (angling in). Similarly, it describes batting features such as response (outside edge) pointing to the batsman’s imperfection. The word false shot emphasizes the imperfection and points to batsman’s weakness against good length and angling in delivery.

In Section 2-3, we summarize two of our recently published work, which utilizes text commentaries to mine strengths and weaknesses of cricket players.

## 2 Learning Strengths and Weaknesses

In this work [1], we have extracted the fine-grained information about each player from the text commentaries and represented them using domain-specific features. We have provided computationally feasible definitions of strength rule and weakness rule and propose to use a dimensionality reduction method, namely Correspondence Analysis, to obtain semantic relations between bowler and batsman. We have used biplots to plot these relations and extracted human-readable rules. Fig. 1 is a biplot for batsman Steve Smith and the inferred strength rule is: *“Smith attacks deliveries that are slow or short-pitched or bowled on middle stump”*.

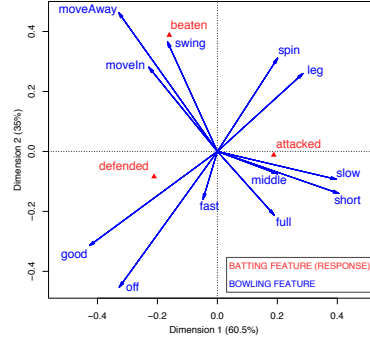


Fig. 1. Smith's response biplot.

## 3 Temporal Changes in Strengths and Weaknesses

In this work [2], we have mined the temporal changes in the obtained strength/weakness rules. We have constructed a tensor for each player using the extracted domain-specific features from the text commentaries. To obtain the dependencies between batting features, bowling features, and time, we have used Three-way Correspondence Analysis, which internally uses tensor decomposition. We have plotted the extracted temporal changes in a line plot. In Fig. 2, blue line represents the change in strength rule for Steve Smith.

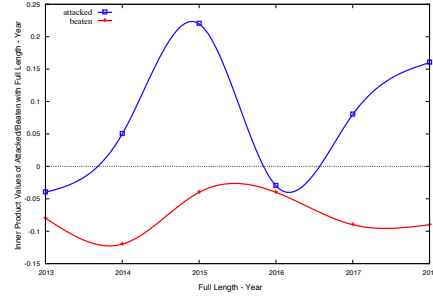


Fig. 2. Smith's strength and weakness on full-length deliveries over the years.

## References

1. Behera, S. R., Agrawal, P., Awekar, A., Vedula, V. S.: Mining Strengths and Weaknesses of Cricket Players Using Short Text Commentary. In: 18th IEEE International Conference On Machine Learning And Applications (ICMLA), pp. 673–679, Boca Raton, FL, USA (2019). <https://doi.org/10.1109/ICMLA.2019.00122>
2. Behera, S. R., Vedula, V. S.: Mining Temporal Changes in Strengths and Weaknesses of Cricket Players Using Tensor Decomposition. In: 28th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), Bruges, Belgium (2020).