

# Using ball-by-ball text commentary data to learn weakness and strength rules of cricket players.

## Batsman's Kryptonite: Learning Weakness and Strength Rules of Cricket Players using Association Rule Mining

👤 Swarup Ranjan Behera and V. Vijay Saradhi, Indian Institute of Technology Guwahati, India

### BACKGROUND

- Knowledge of strengths and weaknesses of players is the key for team selection and strategy planning in any team sport such as Cricket<sup>1</sup>
- Computationally, this problem is mostly unexplored
- Existing methods focus only on aggregate and macroscopic statistics that ignore many details

### OBJECTIVE

- To identify the strength and weakness rules of cricket players

### CHALLENGES

- Which data?
- What constitutes a rule?
- Which computational method?

### APPROACH

- Data
  - Publicly available cricket text commentary data<sup>2</sup>
- Rule (bowling feat  $\Rightarrow$  batting feat)
  - Weakness rule (bat feat: beaten)
  - Strength rule (bat feat: attacked)
- Association rule mining
  - Apriori algorithm<sup>3</sup>

### RESULTS:

#### Strength Rules of batsman Steve Smith:

- {short length, slow}  $\Rightarrow$  {attacked}
- {leg stump}  $\Rightarrow$  {attacked}
- Steve Smith *attacks* the *leg stump*, *slow*, and *shot length* deliveries

#### Weakness Rules of batsman Steve Smith:

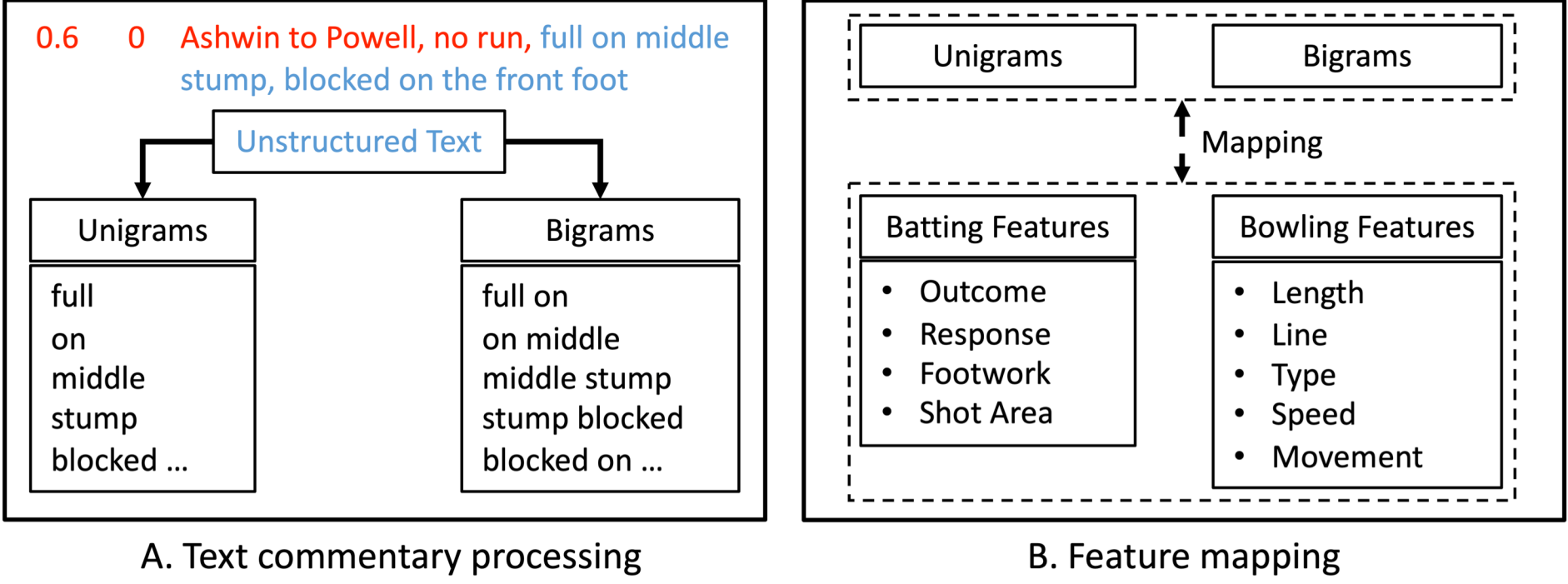
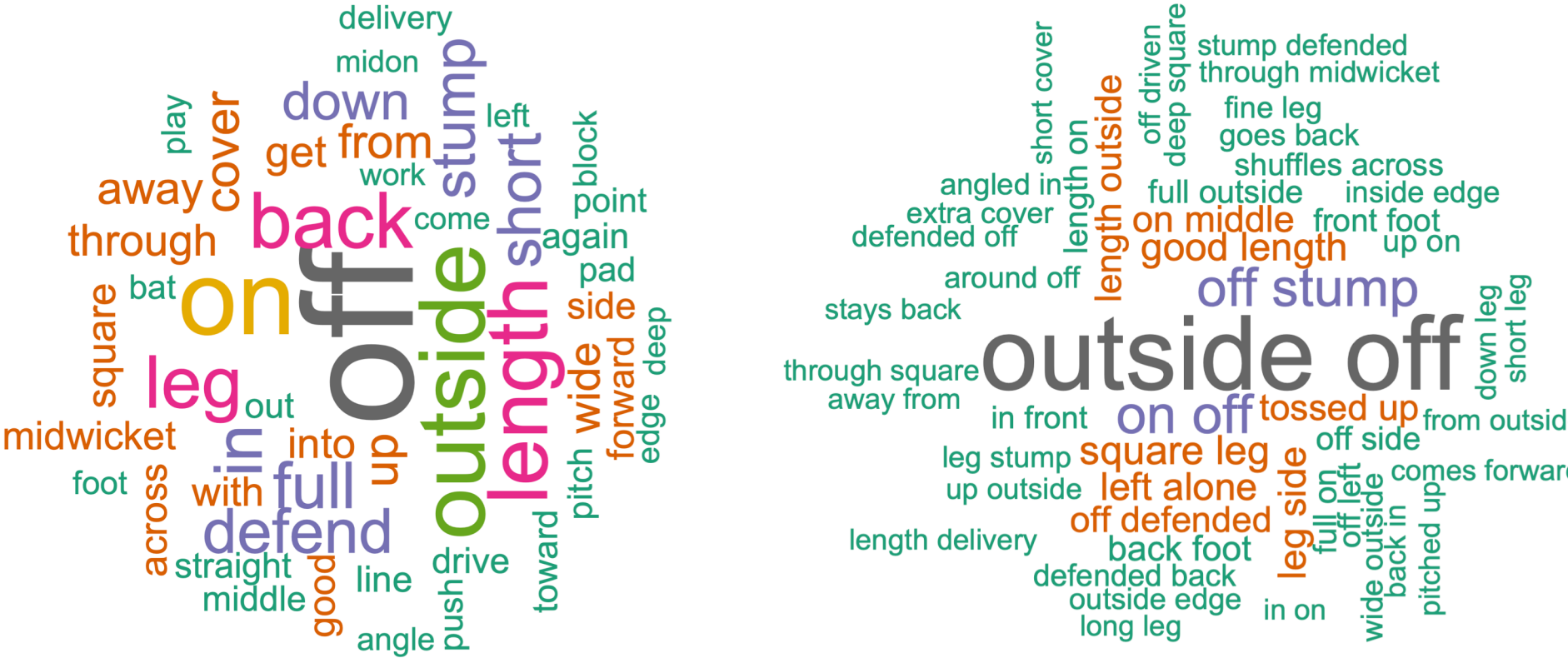
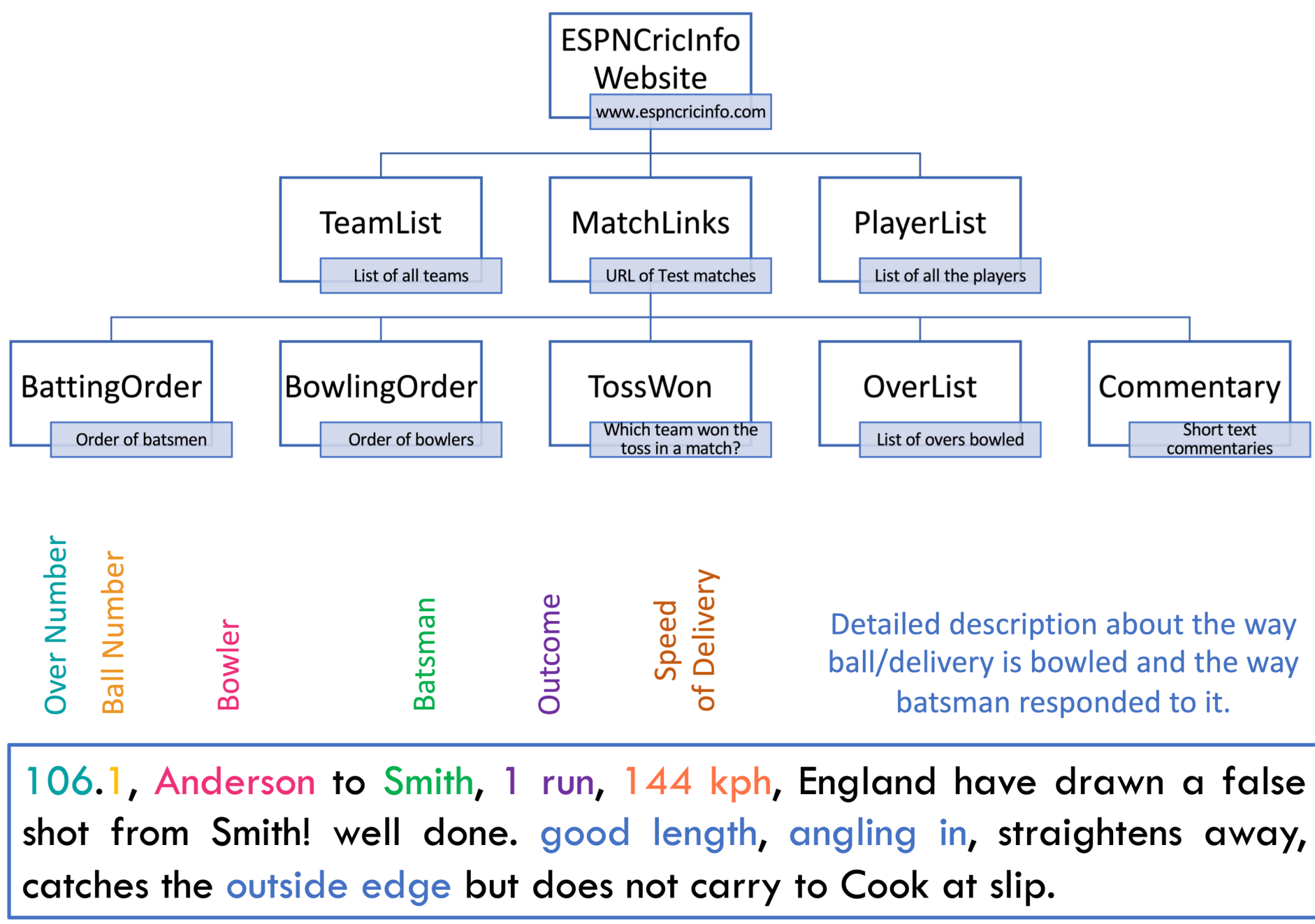
- {fast, good length, swing}  $\Rightarrow$  {beaten}
- Steve Smith gets *beaten* on the *fast*, *good length*, and *swing* deliveries

#### Other Rules of batsman Steve Smith:

- {off stump}  $\Rightarrow$  {defended}
- Steve Smith *defends* deliveries bowled on *off stump* line
- {fast, off stump}  $\Rightarrow$  {0 run}
- Steve Smith scores *zero runs* on *fast* deliveries bowled on *off stump* line

### CONCLUSION

- We presented an application of association rule mining for learning strength and weakness rules
- Rules for more than 500 players are constructed by analyzing text commentaries over one million deliveries for 13 years (2006-2019)
- The constructed rules will help analysts, coaches, and team management in building game strategies



BOWLING FEATURES				BATTING FEATURES			
<b>Outcome</b> (Result of a delivery) 0 (no run) 1 (single) 2 (double) 3 (triple) 4 (four) 5 (five) 6 (six) Out (batsman is out) Not out (batsman is not out)				<b>Response</b> (Response of the batsman on each delivery) Attacked (batsman attacks the ball) Defended (batsman defends the ball) Beaten (batsman is beaten by the ball) Back foot (batsman is on back foot) Front foot (batsman is on front foot)			
<b>Length</b> (How far the ball is bowled) Full (full length) Short (short length) Good (good length) Bad (bad length)				<b>Footwork</b> (How the batsman plays the ball) Square (square bat) Square leg (square leg bat) Fine leg (fine leg bat) Off (off stump) On (on stump)			
<b>Speed</b> (Speed of the ball) Fast (fast delivery) Slow (slow delivery)				<b>Line</b> (Line of the ball) Off (off stump) On (on stump) Good (good length) Bad (bad length)			
<b>Type</b> (Type of the delivery) Spin (spin delivery) Swing (swing delivery) Straight (straight delivery)				<b>Shot Area</b> (Area where the ball is played) Square (square bat) Square leg (square leg bat) Fine leg (fine leg bat) Off (off stump) On (on stump)			
<b>Movement</b> (Movement of the ball) Move-in (move in) Move-out (move out)				<b>Shot Area</b> (Area where the ball is played) Square (square bat) Square leg (square leg bat) Fine leg (fine leg bat) Off (off stump) On (on stump)			

Association Rule (A $\Rightarrow$ B)	Support(%)	Confidence(%)	Lift
{shortlength, slow} $\Rightarrow$ {attacked}	2.6	72.1	1.7
{legstump} $\Rightarrow$ {attacked}	2.5	60.1	1.4
{fast, goodlength, swing} $\Rightarrow$ {beaten}	0.1	37.9	3.9
{off stump} $\Rightarrow$ {defended}	26.8	59.8	1.3
{fast, short length} $\Rightarrow$ {back foot}	9.2	91.4	1.9
{full length, off stump} $\Rightarrow$ {front foot}	8.5	81.9	1.6
{fast, off stump} $\Rightarrow$ {0 run}	22.8	82.2	1.2
{fast, off stump} $\Rightarrow$ {square off}	11.2	50.4	1.4

1. What Is Cricket? (<https://www.icc-cricket.com/video/what-is-cricket>)  
2. EspnCricInfo Website (<https://www.espninfo.com>)  
3. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. In Proc. of the ACM SIGMOD Conference on Management of Data, 22(2), (pp. 207-216).

