Predicting the outcome of ODI Cricket matches

Ganesh Bhambarkar 153050072
Samiran Roy 153050091
Sushant Shambharkar 153050081
Surendra Singh Rao 153050069

February 2016

1 Description

Our task is to investigate to what degree we can predict the outcome of cricket matches, specifically ODI Matches. Given the popularity of this multi-million dollar industry, there is a strong incentive for match fixing and underground betting. The complex rules surrounding the game, along with the numerous parameters affecting the game, including but not limited to cricketing skills and performances, match venues, toss, weather conditions present significant challenges for accurate prediction. This problem has been well investigated for games like basketball and soccer, but yet to be researched for cricket.

2 Dataset

We will be using the dataset from `cricsheet.org`, which provides ball by ball data on 1,164 One-day internationals, in YAML format, others contain all of the matches, and the others certain sub-sets of matches like Test matches, IPL matches etc. They also provide (as an experiment) all T20 internationals, and all matches in CSV format. Below is the listing of the data grouped by types of matches, or you can see the full set of downloads.[1]

3 Description of the method

The task can be divided into three stages:

1) Feature engineering will play a major role in our project. Parsing the YAML file, and extracting relevant features will will be the first task. We will try to extract relevant features such as:

- Winning percentage against teams
- Scoring rates and dismissal probabilities

- History of games at that venue
- Conditions (pitch, ground size, weather)
- Player weaknesses

We will be using other datasets to fill the gaps in our data.

2) We will use techniques like Pearson correlation, recursive feature elimination and random forests for feature selection.

3) Building a machine learning model for predicting the outcome of a cricket ODI match.

We will be studying the cites papers: [3], [4], [2]

4 Plan to submit for 21 March submission

Our first plan is to write code to transform the ball by ball data (YAML) file to structured data. We will submit steps 1 and 2 in the 21 March submission.

5 Plan to submit for the final submission

We plan to submit the whole data transformer, learning models and our results in the final submission.

References

[1] Cricksheet, 2016.

[2] Amal Kaluarachchi and Aparna S Varde. Cricai: A classification based tool to predict the outcome in odi cricket. In *Information and Automation for Sustainability (ICIAFs), 2010 5th International Conference on*, pages 250–255. IEEE, 2010.

[3] Stylianos Kampakis and William Thomas. Using machine learning to predict the outcome of english county twenty over cricket matches. *arXiv preprint arXiv:1511.05837*, 2015.

[4] Vignesh Veppur Sankaranarayanan, Junaed Sattar, and Laks VS Lakshmanan. Auto-play: A data mining approach to odi cricket simulation and prediction. In *SDM*, pages 1064–1072. SIAM, 2014.