An Alternative Machine Learning approach to DuckWorth-Lewis

Authors: Abhiram Eswaran, Akul Swamy

Problem: In case of rain interruption in a limited overs cricket match, the adjustment to the score is made using the traditional D/L method. This takes into consideration, only wickets and overs remaining. However, other factors such as venue, momentum, powerplay, etc; also play a role in determining the fair score. Therefore, this is the motivation behind the problem where we try to perform a more accurate and fair score adjustment.

Methodology:

Features: Once we have the data sets in place, we plan to use the following features for our prediction model.

1. Powerplay overs left.
2. Non powerplay overs left.
3. Wickets left.
4. Average 1st and 2nd innings score at the venue.

Models: We plan to start our prediction problem with Linear Regression, followed by Support Vector Regression and then eventually using ensemble methods.

Code Libraries

- For data collection we plan to use BeautifulSoup.[1]
- For structuring data into a convenient format, we use Pandas.
- Libraries for modeling include Sklearn and Numpy.

Hardware Platform: We intend to use our local machines for training and testing purposes. If the scale in not manageable, we intend to use Amazon AWS for higher order computation.

Data Sets:

We plan to scrape data from Cricinfo which maintains data of all international cricket matches from 1971. We intend to use the data from 1994 from when limited overs cricket began. We plan to scrape ball by ball details of each match. The details of the data can be viewed here.

Experiments:

We consider all 3 scenarios where the rain could impact play: Rain interruption- in first innings, in the second innings and rain abandoning play. For each scenario, we plan to induce rain at a point in the game, look at the actual score (*gold label*) at that point (*From the data scraped from ball by ball details)*, predict our score at that point and compare it with the score predicted by Duckworth Lewis.

Our Evaluation measure: We evaluate our model's predictions with the actual score. The degree of closeness to actual score determines the accuracy of the model.

---

[1] Beautiful soup: https://www.crummy.com/software/BeautifulSoup/

Related Work:

Duckworth-Lewis method[1] is the one which is widely accepted in this scenario. It takes into account, wickets and overs remaining as the only resources of a team. This however does not seem correct and might not always result in a fair score adjustment. Also, with the advent of T20 cricket and power plays, larger totals have been chased down with ease over the recent past in International cricket. The D/L methods fails to take this "acceleration" into account. We address this by considering power play overs as a feature in our model. Also, momentum and venue play key roles in determining the final score and we wish to investigate this by employing ML algorithms to make use of these features.

In the work proposed in [2], the model discusses using momentum, venue, player ratings and player history for predicting scores. While momentum and venues have a positive impact, modeling on player history and player rating impacts the performance of the model negatively since "form" of a player is not quantifiable. We address these problems in [2], by considering just the average score in the venue as a feature.

Collaboration Plan:

The project involves 4 main phases,

1. Data Collection
2. Feature Engineering
3. Prediction Modeling
4. Testing and Evaluation.

We plan to distribute ourselves with 2 components each. With Abhiram taking up data collection and prediction modeling and Akul doing Feature Engineering and Testing with Evaluation.

References:

1. DuckWorth-Lewis Method:
    a. http://static.espncricinfo.com/db/ABOUT_CRICKET/RAIN_RULES/DUCKWORTH_LEWIS.html
    b. https://en.wikipedia.org/wiki/Duckworth%E2%80%93Lewis_method
2. A Learning Algorithm for Prediction in the game of cricket.
    https://people.ucsc.edu/~praman1/static/pub/ML_Project_CS7641.pdf
3. Cricinfo Data: http://bit.ly/2nGXGyZ
4. Using machine learning to predict outcome of Twenty over cricket matches
    https://arxiv.org/pdf/1511.05837.pdf
5. Cricket Wiki: https://www.reddit.com/r/Cricket/wiki/cricket_data