Research Papers

Cricket Squad Analysis Using Multiple Random Forest Regression

IEEE Xplore Link: https://ieeexplore.ieee.org/document/8987367
Sci-Hub Link: https://sci-hub.se/10.1109/ICAIT47043.2019.8987367
Pages: 5

Summary:

Abstract

- Performance analysis is required for a balanced squad
- Different tours, stadiums demand different squads
- To determine the players in these squads -
    - Experience
    - How the player will play under those conditions + other attributes
- The paper uses multiple random forest regression to which will be used to predict the value of the attributes that have to be considered for a batsman/bowler
- The model uses a past performance of the player with against the same opponent to train
- A rank wise list is produced for the team at the end for that particular tour

I. Introduction

- Squad usually contains 15 player chosen based on strengths/weakness
- They are the set of players best suited for the conditions
- ML is used to generate this squad
- The opposition + venue are considered
- *Regression* - Prediction of continuous value for batsman + bowler in a squad
- Dependant variable
    - Batsman - Runs scored
    - Bowler - Number of wickets + Economy rate
- Independant variable (am not sure)
    - Venue, opposition, innings
- Patterns in the data are used for player evaluation
- Match is split into 3 parts
    - Power play
    - Middle
    - Death over
- Multiple algorithms are used to find the attributes in the 3 segments of the game
- *Compatibility between different combinations of the attributes was checked using Naïve Bayes algorithm and Support Vector Machine algorithm*
- Associative rule mining is used for selecting the players for International Indian Team

- *Sentiment analysis was used to create a team for IPL based on training data from Twitter tweets*
- Non-dominated Genetic Sorting Algorithm was used to create an ODI cricket team
    - **Batsman - Batting average**
    - **Bowler - Bowling average**
    - Other labels - fitness combined with strengths
- Performance of players were predicted using **Multiclass SVM, Naïve Bayes, Decision Tree classifiers and Random Forest (Random Forest had highest accuracy)**
- Metrics considered for this - *Consistency, Form, Opposition, Venue*
- **Analytic Hierarchy Process** was used to assign weights
- Data was taken from **Cricinfo + CricBuzz**
- Models for classification whether the player should/should not be a part of the team - **KNN, SVM, Naïve Bayes, and MLP Classifiers** were used
- **Combined Bowling Rate** was used to evaluate bowling performance
- CBR is the combination of economy rate, bowling average and bowler's strike rate is used to calculate the bowler performance
- To predict the score Multiple Variable Linear Regression and Logistic Regression was used
- To predict the team that won the match, Random Forest algorithm was employed
- Genetic algorithms were used to generate 11 players based on the 10 indices for ranking from IPL + T20

II. Problem Description

- Once the algorithm was chosen - it was optimised
- Unpredictable factors - injury + past injury to have detrimental impact on future performance

III. Proposed System/Methodology

- *The data was arranged randomly to ensure that the model is trained accurately*
- The random arrangement may cause underfitting
- **RMSE** - Root Mean Square Error for accuracy
- *Model with the lowest RMSE is considered for training*

VI. Performance Evaluation Parameter

- For optimizing the model the best attributes - that contribute the most are considered
    - *These attributes are considered based on p-values*
    - p-value < 0.05 are considered
    - p-value > 0.05 are discarded

VII. Experimental Setup

- Anaconda distribution
- Train - 80%, Test - 20%

VIII. Conclusion and Future Scope

- The model can be run in the future without having to worry about the formats of the data that have to be considered

- Improve the model by predicting the metrics of a player against a particular bowler/batsman

Pressure Index in Cricket

- Traditional measures used to evaluate individual player performances in one-day cricket include batting and bowling averages, strike and economy rates.
- Various researchers have proposed different measures for assessing batting and bowling performances, but it is acknowledged that these measures have limitations in capturing the true abilities of players.
- Relying solely on averages for evaluating cricketers can lead to skewed and unreliable results.
- Efforts have been made to develop fairer measures of player performance in one-day cricket, but statistics are often presented with verbal context, considering factors like the stage of the innings a player usually bats or bowls, which significantly impacts their opportunities.
- Traditional measures don't allow direct comparison of batting and bowling skills due to incompatible scales.
- The feeling of which team is winning in a T20 match is influenced by the general appreciation of the game and specific match situations.
- A new measure called the "Pressure Index" has been developed to quantify the pressure experienced by a team or a batsman during a T20 match.
- The Pressure Index considers variables such as runs scored, runs left, wickets, balls faced, and balls left to assess the pressure under which a player is performing.
- Runs scored under higher pressure are valued more than those scored under low-pressure situations.
- The Pressure Index allows for a more accurate measurement of a batsman's actual performance, and similar measures can be developed for bowlers.
- Pressure Index = $\left(\frac{Req.\ Rate}{Initial\ Run\ Rate} * 100\right) + \left(\left(\frac{Wicket\ Weight}{180} * Target\right) * \left(\frac{Total\ Balls - Balls\ Faced}{Total\ Balls}\right) * \left(\frac{Target - Runs\ Scored}{Target}\right)\right)$
- For an evenly paced match the pressure index would be exactly 100.
- The number ranges from 0 to some high value above 100.