Fielding-based Ranking of Players in One-Day Cricket

Naimat Ullah*
18030048@lums.edu.pk
Lahore University of Management Sciences
Lahore, Pakistan

Uzair Ghauri*
18030002@lums.edu.pk
Lahore University of Management Sciences
Lahore, Pakistan

Asim Karim
akarim@lums.edu.pk
Lahore University of Management Sciences
Lahore, Pakistan

ABSTRACT

Cricket, a popular sport played by more than a hundred nations, is the second most enjoyed sport after soccer. There has been extensive research in the area of cricket, specifically with regard to simulation and outcome prediction. Most work in this field is based on statistics which prove to be useful for extracting positive contributions of players but often fails to incorporate the negative impact of individual player towards their team performance. Experts and coaches manually observe the playersâĂŹ movements to assess their overall performance. This type of subjective assessment can only be extracted from text data of a match. Having rich text data in this domain can provide excellent results for measuring performance. We attempt to address two problems related to text commentary of matches. First, we propose an approach to generate *Domain Specific(DS)* word embeddings. Secondly, using these *Domain Specific* word embeddings, we aim to extract playersâĂŹ hidden contributions by performing sentiment analysis. Our work suggests that the rela- tive team strength between the competing teams forms a distinctive feature for predicting the winner. Modeling the team strength boils down to modeling individual playersâĂŹ batting, bowling and fielding performances which will form the basis of our approach. Fielding being a very critical aspect for a match outcome, is totally subjective. For a good team, fielding is equally important as batting and bowling. Therefore, we present an approach to qualitatively assess the performance of individual fielders. Combining this approach with stats of batting and bowling can be very helpful for team selection.

KEYWORDS

Sport analytics, word embeddings, sentiment classification

ACM Reference format:
Naimat Ullah, Uzair Ghauri, and Asim Karim. 2019. Fielding-based Ranking of Players in One-Day Cricket. In *Proceedings of*
, 10.1145/1122445.1122456, 4 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Both authors contributed equally to this research.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Primarily played in the Commonwealth member countries , cricket has grown its following across all continents. It has the second largest viewership by population for any sport, next only to soccer, and generates an extremely passionate following among the supporters. There is huge commercial interest in strategic planning for ensuring victory and in game outcome prediction. This has motivated thorough and methodical analysis of individual and team performances, as well as prediction of future games, across all formats of the game. Currently, team strategists rely on a combination of personal experience, team constitution and seat of the pants "cricketing sense" for making instantaneous strategic decisions. Inherently, the methodology employed by human experts is to extract and leverage important information from both past and current game statistics. However, to our knowledge, the underlying science behind this has not been clearly articulated.

Team composition based on players performance plays an essential role in match outcome. A team composed of high rated players often perform well. But team selection and players rating which is currently being used is insufficient. Players rating is a complex decision that goes beyond simply taking into account few performances by the player. The players' unique distinct features comprised of batting averages, run rate, strike rate (for batsmen), bowling averages, economy (for bowlers) as well as their hidden contribution towards team progress. Currently used systems for rating calculation only consider the statistical measures and based on that, provide different rankings for each category except fielding. In this work, we aim to include hidden contribution while rating a player, this latent impact can be gathered from their fielding performance. This type of hidden performance can only be calculated by manual observation or expert opinions about match. As we have text data of commentators remarks, we aim to use this data to extract useful information.

In this work we are classifying text commentary about fielding effort of players, thus ranking them on the fielding effort. Sentiment classification aims to predict the sentiment polarity, such as "positive" or "negative", over a piece of review. It has been a long-standing research topic because of its importance for many applications such as social media analysis, ecommerce, and marketing. Deep learning has brought in progress in various NLP tasks, including sentiment classification.

Various techniques have been used for sentiment classification through different classifiers, some researchers used feed forward or recurrent neural networks. These works directly take the word embeddings pre-trained for general purpose as initial word representations and may conduct fine tuning in the training process.

Some other researchers look into the problem of learning task specific word embeddings for sentiment classification aiming at solving some limitations of applying general pre-trained word embeddings. For example, [Andrew et al. 2014] develop a neural network model to convey sentiment information in the word embeddings. As a result, the learned embeddings are sentiment-aware and able to distinguish words with similar syntactic context but opposite sentiment polarity, such as the words "good" and "bad". In fact, sentiment information can be easily obtained or derived in large scale from some data sources (e.g., the ratings provided by users), which allows reliable learning of such sentiment-aware embeddings.

Apart from these words (e.g. "good" and "bad") with consistent sentiment polarity in different contexts, the polarity of some sentiment words is domain-sensitive.

For example in cricket, the word 'slip' refers to the field position next to the wicket-keeper. Semantically, it's quite different from it's general meaning which is to loose one's footing. The performance of generic word embeddings in such applications is limited, since word embeddings pre-trained on generic corpora do not capture domain specific semantics/knowledge, while embeddings learned on small data sets are of low quality.

The main problem revolves around sentiment classification. Given raw text, we need to label the effort of the player as good, bad or neutral. In fact, for more accurate results, we need to quantify how good or bad the effort was.

2 RELATED WORK

2.1 Sports Analytics

The problem of match outcome prediction has been studied extensively in the context of basketball and soccer. [Bhandari et al. 1997] developed the Advanced Scout system for discovering interesting patterns from basketball games, which has is now used by the NBA teams.

One of the earliest and pioneering works in cricket was by Duckworth and Lewis [F. and A. 1998] where they introduce the Duckworth-Lewis or D-L method, which allows fair adjustment of scores in proportion to the time lost due to match interruption (often due to adverse weather conditions such as rain, poor visibility etc.). This proposal has been adopted by the International Cricket Council (ICC) as a means to reset targets in matches where time is lost due to match interruptions. The method proposed in [F. and A. 1998], and subsequently adapted by [McHale and Asif 2013], for capturing the resources of a team during the progression of a match has found independent use in subsequent work in cricket modeling and mining [McHale and Asif 2013][M. and S. 2006].

Lewis [Lewis 2005], Lemmer [Lemmer. 2008], Alsopp and Clarke [P. and Stephen 2004] , and Beaudoin [D. 2003] develop new performance measures to rate teams and to find the most valuable players. Raj and Padma [Raj and Padma 2013] analyze the Indian cricket team's One-Day International (ODI) match data and mine association rules from a set of features, namely toss, home or away game, batting first or second and game outcome. Kaluarachchi and Varde [Kaluarachchi and Varde 2010] employ both association rules and naive Bayes classifier and analyze the factors contributing to a win, also taking day/day-night game into account. Both approaches use a very limited subset of high-level features to analyze the factors contributing to victory.[Harsha et al. 2017] asses the impact of fielding by introducing a metric of expected runs saved due to fielding this work also lack the use of text data.

2.2 Word Embeddings and Sentiment Analysis

As we aim to classify fielders' performance based on positive and negative contribution by sentiment analysis of text data. Traditional vector space models encode individual words using the one-hot representation, namely, a high-dimensional vector with all zeroes except in one component corresponding to that word [Baeza-Yates and Ribeiro-Neto 1999] . Such representations suffer from the curse of dimensionality, as there are many components in these vectors due to the vocabulary size. Another drawback is that semantic relatedness of words cannot be modeled using such representations. Later,Mikolov [Mikolov et al. 2013] propose two methods that are considerably more efficient, namely skip-gram and CBOW. This work has made it possible to learn word embeddings from large data commentary sets, which has led to the current popularity of word embeddings. Glove and Word2Vec are two modern neural probabilistic language models which can be applied in many NLP tasks, like named entity recognition[Joseph Turian and Bengio 2010], document classification[Bei Shi and Lai 2017], word sense disambiguation and parsing.

Sentiment classification has been a longstanding research topic [Liu 2012] [Pang and Lee 2008] [Peng et al. 2017] Given a review, the task aims at predicting the sentiment polarity on the sentence level [Yoon 2014] or the aspect level [Peng et al. 2017]. Supervised learning algorithms have been widely used in sentiment classification [Bo et al. 2002]. People usually use different expressions of sentiment semantics in different domains. Due to the mismatch between domain specific words, a sentiment classifier trained in one domain may not work well when it is directly applied to other domains.

Some researchers have proposed some methods to learn task-specific word embeddings for sentiment classification. Compared to above works, our model focus on learning domain specific and domain-common embeddings to our task of player contribution analysis.

3 DATASET

Dataset consists of all the One-Day Internationals (ODIs) played between years 2005-2019. Original dataset consists of JSON files containing ball-by-ball commentary. Alongside commentary, bowler's id, batsman's id and runs scored on that ball is also available. This dataset is used to generate new dataset for our problem to enhance efficiency.

4 PROBLEM FORMULATION

International Cricket Council (ICC) regularly publishes player rankings for batsmen, bowlers and all-rounders. These rankings are purely based on statistics which include batting average, bowling average, batting strike rate, bowling strike rate, etc. and are usually a good measure for evaluating players' performance.

Fielding, which is the 3rd most important element of cricket, has no ranking system. This is because the only statistical measure available to evaluate fielding caliber of a player is the number of catches they have caught and dropped. Most of the fielding includes ground fielding which cannot be described quantitatively; and is very subjective in nature.

Given ball-by-ball commentary, our goal is to quantify the effort of fielder (if involved) on each ball. This information will then be used to devise a fielding ranking system.

5 TECHNIQUE

The workflow of our methodology is as follows:

(1) Very first step is to separate out comments which mention any fielding effort. A comment may contain information about batsman, bowler, umpire or any incident happened on that ball. Each comment is split into shorter sentences/phrases based on specific rules. For example, splitting on the basis of punctuation marks (full stop, exclamation mark) and keywords like 'and', 'so', etc.
(2) Develop domain specific word embeddings using Word2Vec [Mikolov et al. 2013]. These embeddings are used in classification as well as sentiment analysis.
(3) Train a classifier on labelled data. Labels are of four types: batting, bowling, fielding and other.
(4) The trained classifier is then used to predict labels of more data.
(5) Perform sentiment analysis on comments labelled as 'fielder'. Quantify fielding effort based on the output of sentiment analysis.
(6) Use results to generate player rankings.

6 EXPERIMENTATION

6.1 Classification

We experimented with multiple classifiers to separate out comments related to a fielding effort. For each classifier, input is given in TF-IDF format. Results are summarized in Table 1. Multinominal Naive Bayes, with the accuracy of 86% for the fielding class, gave the best performance. Other classifiers include Stochastic Gradient Descent (SGD) and Linear SVM with 83.1% and 79.8% accuracies respectively.

Classifier	Overall accuracy	Fielding accuracy
MNB	72.2%	86%
SGD	74.9%	83.1%
Linear SVM	73.9%	79.8%

Table 1: Classification

6.2 Sentiment Analysis

Different lexicons for sentiment analysis

6.2.1 AFINN. is a list of English words rated for valence with an integer. between minus five (negative) and plus five (positive). Score of each word in the sentence/phrase is summed up to predict whether it belongs to positive, negative or neutral class.

Since AFINN score of each word is independent of the other, Figure 1 shows that its accuracy in our case is around 40

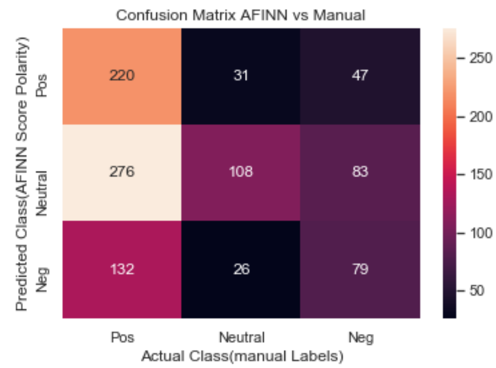

Figure 1: AFINN Score

6.2.2 SenticNet. consists of a set of tools and techniques for sentiment analysis combining commonsense reasoning, psychology, linguistics, and machine learning. SenticNet provides a set of semantics, sentics, and polarity associated with 100,000 natural language concepts. Sentics are emotion categorization values expressed in terms of four affective dimensions (Pleasantness, Attention, Sensitivity, and Aptitude) and polarity is floating number between -1 and +1 (where -1 is extreme negativity and +1 is extreme positivity).

Figure 2 shows the confusion matrix for the results of SenticNet. It can be clearly seen that, in our case, SenticNet is biased towards positive class with overall accuracy of less than 50%.

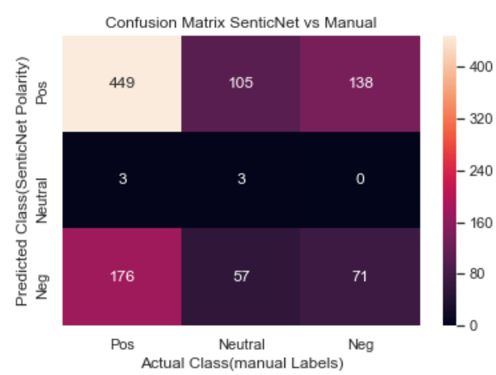

Figure 2: SenticNet

6.2.3 Sentiword Net. is a lexical resource in which each word is associated to three numerical scores Obj(s), Pos(s) and Neg(s), describing how objective, positive, and negative the word is.

SentiWordNet works pretty well for simple sentiment analysis of text. But when it comes to particular domains, as depicted from Figure 3, its performance is worse than SenticNet.

6.2.4 Relaxed Accuracy. The results of common sentiment analysis techniques were not only poor (hardly touching 50% accuracy) but also insufficient to devise a ranking system. Most of the tools were only able to classify fielding effort as positive, negative or

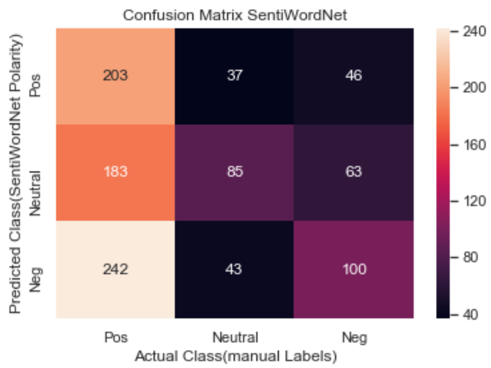

Figure 3: Sentiword Net

neutral but for a reasonable ranking system, we also needed to quantify the effort. For this purpose, data was labelled in 11 classes from -5 to +5 with -5 representing the worst fielding effort and +5 the best.

Before presenting results, we introduce a term called 'Relaxed Accuracy'. It is the measure we used to evaluate the performance of different classifiers. Predicted class is considered to be true if it lies within ±1 range of the true label. For example, if the true label was +3 and the classifier predicted it to be +2 or +4, then also it is considered to be the correct prediction.

6.2.5 Traditional Classifiers. Results of multiple traditional classifiers are summarized in Table X. SVM, with the relaxed accuracy of 78.4% gave the best performance. Other classifiers include Multinomial Naive Bayes (MNB) and Logistic Regression with relaxed accuracies of 73 and 76 percent respectively.

6.2.6 Neural Networks. Simple feed-forward neural network with pretrained GloVe embeddings gave the relaxed accuracy of 64.1%. This was improved to 71.7% using LSTM with domain-specific embeddings. Although the results of neural networks are not better than the traditional classifiers but we believe that these can be improved by training the network on more number of examples.

Classifier	Overall accuracy	Fielding accuracy
MNB	32.2%	73%
SGD	42.5%	78.4%
Logistic Regression	43.1%	76%

Table 2: Sentiment Analysis: Traditional Classifiers

7 CONCLUSION

Rankings were calculated for Pakistan and England team players for matches from 2015-18. Since there is no benchmark to evaluate the goodness of such ranking system, our subjective evaluation suggests that our approach has considerable potential to become a useful tool not only for cricket but also for other popular sports like football and basketball.

Future extensions include

- smarter sentence splitting approaches
- use other information like Umpires' names and stadium name
- handle substitute fielder cases
- check for effort of more than one fielders on a ball
- re-use same approach for finding hidden contribution of batsmen and bowlers

REFERENCES

Quanbeck Andrew, Socher Richard, and Manning Christopher. 2014. Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), 1532–1543.

TRicardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. Modern Information Retrieval. *ACM press New York* 463 (1999).

Shoaib Jameel Steven Schockaert Bei Shi, Wai Lam and Kwun Ping Lai. 2017. Jointly learning word embeddings and latent topics. *In Proceedings of the $40_t h$ International ACM SIGIR Conference on Research and Development in Information Retrieval.* (2017), 375–284.

Inderpal Bhandari, Edward Colet, Jennifer Parker, Zachary Pines, Rajiv Pratap, and Krishnakumar Ramanujam. 1997. Advanced Scout: Data Mining and Knowledge Discovery in NBA Data. *Data Mining and Knowledge Discovery* 1, 1 (01 Mar 1997), 121–125. https://doi.org/10.1023/A:1009782106822

Pang Bo, Lee Lillian, and Vaithyanathan Shivakumar. 2002. Thumbs up?: sentiment classification using machine learning techniques. *In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing.* (2002), 79–86.

Beaudoin D. 2003. The Best Batsmen and Bowlers in One-Day Cricket. *PhD thesis, Simon Fraser University* (2003).

Duckworth F., C. and Lewis A., J. 1998. A fair method for resetting the target in interrupted one-day cricket matches. *The Journal of the Operational Research Society* 49, 3 (1998), 220–227.

Perera Harsha, Davis Jack, and Swartz Tim. 2017. Assessing the impact of fielding in Twenty20 cricket. *Journal of the Operational Research Society* 69, 8 (2017), 1135–1343. https://doi.org/10.1080/01605682.2017.1398204

Lev Ratinov Joseph Turian and Yoshua Bengio. 2010. Word representations: a simple and general method for semi-supervised learning. *In Proceedings of the $48_t h$ Annual Meeting of the Association for Computational Linguistics* (2010), 284–294.

A. Kaluarachchi and A. Varde. 2010. CricAI:A classification based tool to predict the outcome in ODI cricket. *In 5th International Conference on Information and Automation for Sustainability* 49, 3 (2010), 250–255.

H. H. Lemmer. 2008. An analysis of playersâĂŹ performances in the first cricket Twenty20 World Cup series. *South African Journal for Research in Sport, Physical Education and Recreation* 30, 2 (2008), 71–77.

A. J. Lewis. 2005. Towards fairer measures of player performance in one-day cricket. *The Journal of the Operational Research Society* 56, 7 (2005), 804–815.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5, 1 (2012), 1–167.

Bailey M. and Clarke S., R. 2006. Predicting the match outcome in one-day international cricket matches, while the game is in progress. *Journal of sports Science and Medicine* 5, 4 (2006), 480–487.

I. G. McHale and M. Asif. 2013. A modified DuckworthLewis method for adjusting targets in interrupted limited overs cricket. *European Journal of Operational Research* 225, 2 (March 2013), 353–362. https://doi.org/10.1016/j.ejor.2012.09.036

Tomas Mikolov, Ilya Sutskever, Chen Kai, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *In Advances in neural information processing systems* (2013), 3111–3119.

Allsopp P., E. and Clarke Stephen, R. 2004. Assessing the impact of fielding in Twenty20 cricket. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 167, 4 (2004), 657–667. https://doi.org/10.1111/j.1467-985x.2004.00505.x

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval. *Synthesis Lectures on Human Language Technologies* 2, (1-2) (2008), 1–135.

Chen Peng, Sun Zhongqian, Bing Lidong, and Yang Wei. 2017. Recurrent attention network on memory for aspect sentiment analysis. *In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (2017), 452–461.

K. Raj and P. Padma. 2013. Application of association rule mining: A case study on team India. *International Conference on Computer Communication and Informatics (ICCCI)* (2013), 1–6.

Kim Yoon. 2014. Convolutional neural networks for sentence classification. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (2014), 1746–1751.