PREDICTING THE OUTCOME OF THE CRICKET MATCHES USING MACHINE LEARNING TECHNIQUES

Kausik M. , Siyamalan M.

Department of Computer Science, Faculty of Science , University of Jaffna

kausik219@gmail.com, siyam@univ.jfn.ac.lk

SLSWCS Student Workshop on Computer Science

Abstract

This poster proposes two novel approaches for predicting the outcome of cricket matches by modeling the team performance based on the performances of its players in other matches. Our first approach is based on feature encoding, which assumes that there are different categories of players exist and models each team as a composition of player–category relationships. The second approach is based on a shallow Convolutional Neural Network (CNN) architecture, which contains only four layers to learn an end-to-end mapping between the performance of the players and the outcome of matches. Both of our approaches give considerable improvement over the baseline approaches we consider, and our shallow CNN architecture performs better than our proposed feature encoding based approach. We show that the outcome of a match can be predicted with over 70% of accuracy.

Methodology

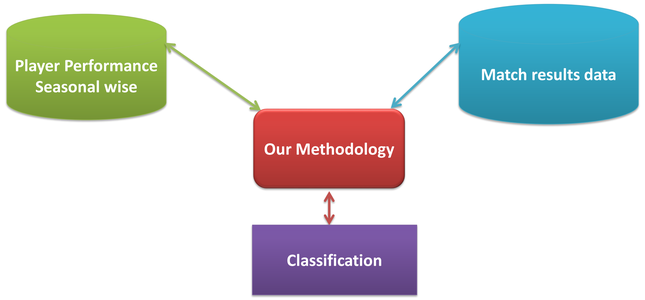
Let $X_{i,t}^{s} \in \mathbb{R}^{d}$ represents the performance of a player P_i from the team T_t at the season s, where $X_{i,t}^{s}$ can be represented by the player P_i's batting, bowling and fielding capability at the season s.

Team T_t^s can be represented as { $X_{1,t}^{s-1}, X_{2,t}^{s-1}$, ... , $X_{11,t}^{s-1}$ }; Here the important thing is that the season should be *s-1*, because the match season is s therefore we need to get the player performance from the previous season to represent the team composition. Then only we can get the recent performance of the player. The outcome is represented as either 0 or 1; if the outcome is 1 then the Team 1 is won against the Team 2(Opponent); if the outcome is 0 then the Team 2 (Opponent) is won against Team 1. We eliminated the matches if the outcome is draw or no result.

Mainly, The following are the two different approaches carried out:

- Feature encoding approach
- CNN approach

Basic idea of the methodology

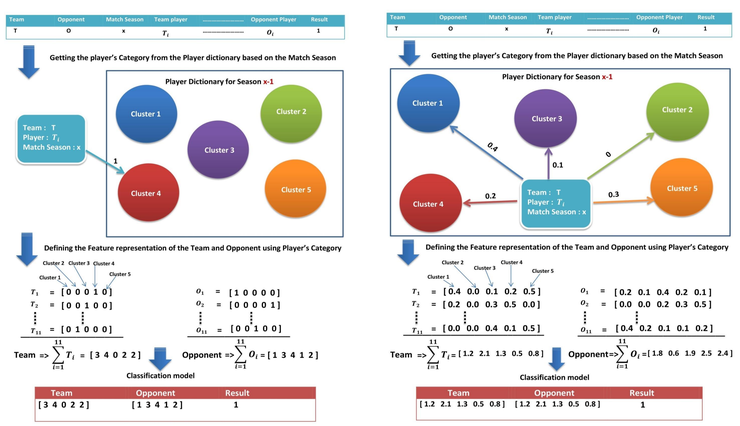


Feature Encoding Approach

We used two types clustering techniques define the player categories

- Hard Assignment model
- Soft Assignment model

We used linear SVM for the classification process with this trained model to get the output.



Hard Assignment model

Soft Assignment model

Convolution Neural Network approach

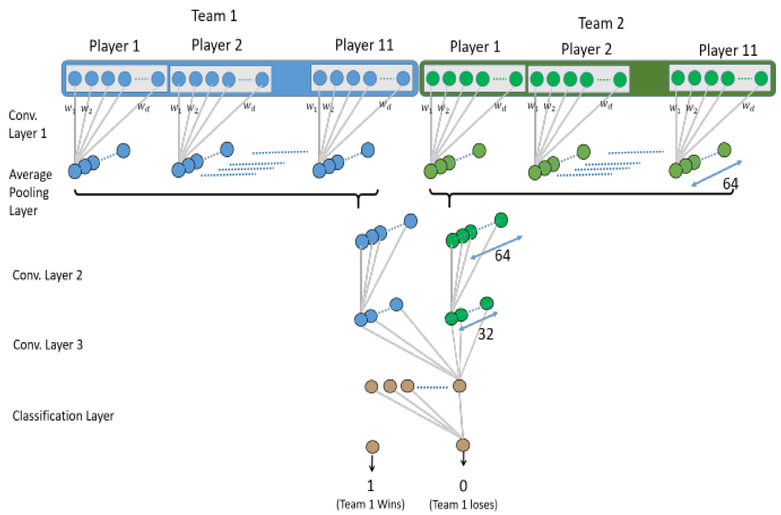


Fig. 1. The proposed CNN architecture. Some connections of the CNN in this figure are not shown to avoid clutter.

TABLE I
PROPOSED CNN ARCHITECTURE. THERE WAS NO PADDING USED IN ALL THE LAYERS.

	Conv. Layer 1	Avg. Pooling	Conv. Layer 2	Conv. Layer 3	Linear classification layer
Input dimension	$d \times 1 \times 1$	$22 \times 1 \times 64$	$2 \times 1 \times 64$	$2 \times 1 \times 32$	$16 \times 1 \times 1$
Output dimension	$22 \times 1 \times 64$	$1 \times 1 \times 64$	$2 \times 1 \times 32$	$1 \times 1 \times 16$	2
Filter size	$d \times 1 \times 1$	$22 \times 1 \times 1$	$1 \times 1 \times 64$	$2 \times 1 \times 32$	$16 \times 1 \times 1$
Stride	d	1	1	1	–
No. of filters	64	–	32	16	–

Dataset

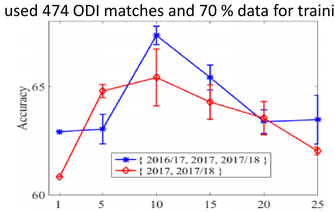
We crawled the website www.espncricinfo.com using our Python wrapper codes to collect the dataset. We were able to collect players information and the details of ODI matches from May 2013 to October 2017. Our dataset contains 2, 581 players and 474 ODI matches in total.

Season	2013	2013/14	2014	2014/15	2015	2015/16	2016	2016/17	2017	2017/18
No. of matches	47	59	35	91	33	90	37	60	51	12

Testing results

We use *Accuracy* as the evaluation measure. For the proposed Feature encoding-based approach, we iterate each experiment 10 times and report the average and standard deviation of accuracy values over these iterations.

In these approaches, we used 474 ODI matches and 70 % data for training and 30 % data for testing.



Effect of dictionary size when tested on different seasons. Vertical bars shows the standard errors.

Testing set (seasons)	N_{tr}	N_{te}	Baseline 1	Baseline 2	FE	CNN	CNN ensemble
{2016/17, 2017, 2017/18/}	634	314	56.68	62.42	67.36 ± 1.02	68.11 ± 1.07	$\mathbf{69.58 \pm 3.33}$
{2017, 2017/18}	708	240	62.02	62.50	65.42 ± 3.21	71.02 ± 3.33	**75.00**

Comparison of different approaches for cricket outcome prediction: baseline 1 and baseline 2 represent the concatenated feature representation and the averaged representation approaches respectively. FE represents the proposed feature encoding approach with a fixed dictionary size of 10. N_{tr} AND N_{te} represent the number of training and testing matches (augmented) respectively.

Conclusion and Discussion

In this poster, we presented two novel approaches for predicting the outcome of Cricket matches. Both of our approaches perform considerably better than the baseline approaches we considered. To the best of our knowledge, we are the first group to apply CNN for Cricket outcome prediction. We showed that the outcome of a Cricket match can be predicted with an accuracy of over 70%. In this work, we modeled each team based on the performance of the players play for that team. We considered features such as number of matches played in each season, number of runs scored, number wickets captured, and so on. However, we haven't considered some of the crucial factors such as the ranking of the players, player's strike rate, etc. These factors along with a larger dataset will be considered in our future work.

References

- Kaluarachchi, Amal and Varde Aparna .S, "CricAI: A classification based tool to predict the outcome in ODI cricket", *2010 Fifth International Conference on Information and Automation for Sustainability*, IEEE 2010.
- Shah, P. , "Predicting Outcome of Live Cricket Match Using Duckworth- Lewis Par Score", *International Journal of Systems Science and Applied Mathematics*, 2017.
- Madan Gopal Jhanwar and Vikram Pudi, "Predicting the outcome of ODI cricket matches: A Team Composition based Approach", *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2016)*, 2016.
- Sankaranarayanan, Vignesh Veppur, Junaed Sattar, and Laks VS Lakshmanan, "Auto-play: A Data Mining Approach to ODI Cricket Simulation and Prediction", SDM. 2014.