

Convolutional Neural Network and Feature Encoding for Predicting the Outcome of Cricket Matches

Siyamalan Manivannan, Mogan Kausik

Department of Computer Science, University of Jaffna, Sri Lanka

Abstract— This paper proposes two novel approaches for predicting the outcome of cricket matches by modelling 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.

Index Terms—Convolutional Neural Networks, Feature Encoding, Winning team prediction in Cricket

I. INTRODUCTION

Cricket is the most popular sport in Sri Lanka. It is played by two teams with eleven players in each side. Each team attempts to score runs, and team with the most runs will be the winner. International Cricket Council recognizes three forms of Cricket mainly based on the scheduled duration of the game. They are, Test Cricket, One Day International (ODI), and Twenty-20. ODI is considered the most popular form of Cricket, one of the reason is it allows day-night format compared to classical day-only form [1].

Outcome prediction of a particular match (which team will win in that particular match) even before the match is scheduled would be very useful in Cricket. If the prediction is based on the players' performance, a particular team could be composed against another team in such a way that the composition leads to victory.

There are many factors which can affect the outcome of a Cricket match, which includes who wins the toss, weather condition, condition of the ground, whether the match is held in a team's home country or not, etc. Other than these external factors, the main internal factor which determines the the winning team is the performance of its players. The performance of each player can be determined based on how well he played in other matches, e.g. his average batting score, how many wickets he captured, how many catches he caught, etc.

Applying machine learning techniques to the field of Cricket is recently gained popularity. There are a number approaches have been proposed in the literature for predicting the outcome of the Cricket matches [2]–[5], and predicting the performance of the players [6], [7]. Since we focus on the outcome prediction of the Cricket matches we limit our literature survey on the methods proposed for outcome prediction.

Various features and classifiers have been considered in the literature for predicting the outcome of a Cricket match. For example, factors such as home-field advantage, coin-toss result, bat-first or second, and day vs day-night effect were widely considered [1], [3], [5], [8]. Choudhury et al. [2] considered performances of the teams to determine the outcome; For each team the number of matches played, the number of matches won, the recent standing of the team, number of times a team reached the semi-final stages of a tournament, etc. have been considered as features to determine each team. However, a team's strength will differ based on the team composition (which players play for that team) and the players' performances, and the outcome of a match does not depend only on the factors such as home-field advantage, coin-toss result, bat-first or second, and day vs day-night effects. Differently from these approaches players' performances were considered to build the strength of the teams in [9] and [10] with various pre-determined parameters and rules. Various classifiers such as ANN [2], Random Forest [3], Support Vector Machines [3], Logistic regression [4], and Decision Tree [1], [5] also have been used for predicting the outcome of the matches.

Differently from these approaches, in this paper we propose two novel approaches, one is based on Feature Encoding, and the other is based on CNN, for predicting the outcome of a Cricket match even before the match is scheduled. Our proposed approaches are based on the past performances of the players in each team. In our first approach, we assume that there are different categories of players, and we model each team as a set of player-category relationships, and then train a linear SVM classifier on these relationships. Our second approach is based on a shallow (four layers) CNN, which is trained in an end-to-end manner to extract discriminative features and to learn a classifier to predict which team will win in a particular match. Although we apply our proposed approaches to ODI matches, we believe that our approaches are general, and hence, can be applied to other matches such