



Information Technology and Quantitative Management (ITQM 2017)

Application of Design of Experiment Method for Sports Results Prediction

Fangyao Liu^{a,1}, Yong Shi^a, Lotfollah Najjar^a

^aUniversity of Nebraska at Omaha, 1110 S. 67th Street, Omaha, NE 68182, USA

Abstract

Sports results prediction is increasingly important than many years ago, many factors can impact the results. Other than prediction accuracy, easy to predict is also significant. This research paper explores three easy to grab factors and uses DOE (Design of Experiment) method to determine the effects on sports result prediction. Another goal is to explore the three factors' importance. This research data modeling is based on the NBA (National Basketball Association) game data. All the data is extracted from an official and public data source of 2015 to 2016 regular season.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Keywords: Sports prediction; Design of Experiment; Predict modeling; Data quality management

1. Introduction

1.1. Sports results prediction

Many sports are performing every day and everywhere. The results prediction is making a difference for a lot of people, websites, organizations, and others throughout the world. Predicting game results has become widely popular among sports fans, especially soccer and basketball fans. However, sports analytics have been historically most active in baseball. The reason being the accessibility of baseball data and a growing analytics community behind it (Haghighat et.). Sport analyze work is the application of the management of structured historical data, associated with predictive analytic models that utilize that data, and the use of information systems to inform decision makers and enable them to help their organizations in gaining a competitive advantage on the field of play (Lindsey, G. R.). The world of sports has experienced an explosion in the use of analytics (Alamar & Mehrotra). Prediction is a process of using current sports data to forecast what will happen in the future. There are many methods and theories to analyze and predict team performance, game results or players. One of the greatest changes in the sports world in the past 20 years has been the use of mathematical methods to analyze performances, recognize trends and patterns, and predict results (Severini, T. A.). And, a lot of work is with the help of data gathering, other than peoples opinion and experience. Kinds of professional software make prediction more efficient. For instance, Paul Bessire, a statistician and veteran sports journalist built the state-of-the-art and unbiased predictor, the prediction machine, which strives to be the most accurate and trusted source for predicting sports outcomes straight-up and against-the-spread.

*Corresponding author. Tel.: +1-402-889-9221 .

E-mail address: fangyaoliu@unomaha.edu.

1.2. Prediction evaluation

The evaluation process is the judgment of the predict results. There are many indicators or criteria can be used as the judgment, such as speed, accuracy, reliability, and others. Among them, accuracy is one of the most widely used indicators. Accuracy measures the ratio of correct predictions to the total number of cases evaluated, no one can deny that the ratio of correct predictions to cases should not be a key metric.

This research paper will use another indicator as prediction assessment. It is easy to grab, which is the input factor accessibility with limited resources. As the improvement of the Internet, software, algorithms and other reasons, the current prediction models are getting increasingly complexity and resource consuming. Many people believe the more investment amounts of resources in algorithm development, or data collection, or data quality framework, the better predict results will be. One existing sports prediction model, more than 20 attributes are involved. It is important to spend some time deciding what are going to be the features vectors (input) of the methods, otherwise, the results will be frustrating.

There are many existing models used for sports game outcome prediction. For example, David Orendorff and Todd Johnson used the Bayesian Logic (BLOG) and Markov Logic Networks (MLNs) to predict the NBA game outcome. Their project considers the task of predicting the winner of professional basketball games based on historical data. After the prediction accuracy of models implemented in the BLOG and MLN frameworks are compared using cross validation for the 2006-2007 season. MLN method got 64% accuracy and 63% accuracy for the BLOG model respectively (Orendorff, D.).

This paper proposed only three factors as the inputs to predict the results. All the three factors data are very easy to collect. And the factors are not to responsible for predicting the specific score results, but, they can predict the "Win or Lose" results. Meaning while the history of specific score data is used as training data. The proposed prediction method may have less accuracy, but easier to grab. From the perspective of data resource accessibility along with time and technology resource availability, the complexity prediction model may not have great inputs. When people want to know one basketball game's result is "win or lose", sometimes, they dont have time to find all the required dozens of data inputs, they may just know some basic data inputs, like their ranking number, they play a home game or not. Basically, this paper is to explore the easy to grab data is enough to predict the sports results or not.

2. DOE

2.1. DOE method

The design of experiments (DOE, DOX, or experimental design) is the design of any task that aims to describe or explain the variation of information under conditions that are hypothesized to reflect the variation (Peter). The goal of DOE is to predict the outcome by introducing a change of the preconditions, which is reflected in a variable called the predictor. This research paper uses the systematic method-DOE, to determine the relationship between factors affecting the sports game and the result of that game. Under the help of some statistical tools and experimentation concepts, the cause-and-effect relationships will be verified. For example, the randomization is used, which is the process of assigning individuals at random to groups or to different groups in an experiment, so that every individual of the population has the same chance of becoming a participant in the study.

The following figure is showing the relationship of the factors and results through DOE methodology (Sundararajan).

- Controllable input factors, or x factors, are those input parameters that can be modified in an experiment or process.
- Uncontrollable input factors are those parameters that cannot be changed.
- Responses, or output measures, are the elements of the process outcome that gave the desired effect.

Download English Version:

<https://daneshyari.com/en/article/6901410>

Download Persian Version:

<https://daneshyari.com/article/6901410>

[Daneshyari.com](https://daneshyari.com)