Outlier Detection and Game Outcome Prediction of NBA Game

Aravind Anantha
aananth3@ncsu.edu

Abinav Pothuganti
apothug@ncsu.edu

Chethan Thipperudrappa
cnanniv@ncsu.edu

Huy Tu
hqtu@ncsu.edu

1 Dataset

The dataset is obtained from basketballreference.com and it has All player season stats include the 2004-05 regular season and playoff stats. NBA draft results available are only through 2004 draft. All-Star games are only through the 2003-04 season. Coaching records do not include 2004-05 season. The data has been split to different files based on the information they hold.

2 Project Idea

NBA is one of the most followed games in the US. Finding the outstanding players can be challenging considering the number of attributes affecting this decision. We propose to work on the available NBA dataset to find the outstanding players. This falls into the category of outliers detection and it is widely used in many crucial applications.

There is a lot of interest towards predicting the outcome of a game based on the past games. We also propose to predict the outcome of a game to the highest accuracy possible by experimenting on a range of models for the NBA statistics data.

3 Related Softwares to be Written

Outliers package in R can be used to detect the outliers from the data set. To build a decision tree, an R library called rpart can be used. Naive Bayes method can be implemented using the naiveBayes function available in R. SVM can be implemented using svm method in R. The software will be predominantly written in R and Matlab could be used if necessary.

4 Relevant Papers to Read

- Efficient Algorithms for Mining Outliers from Large Data Sets (can be accessed through this link: `http://dl.acm.org/citation.cfm?id=335437`)
- The Use of Data Mining for Basketball Matches Outcomes Prediction (can be accessed through this link: `http://ieeexplore.ieee.org.prox.lib.ncsu.edu/document/5647440`)

5 Work Division

- Abinav & Aravind - Work on predicting the outcome of the game.
- Chethan & Huy - Work on Outlier Detection.

6 Midterm Milestone

We will aim to complete the outlier detection part of the problem statement and implement basic versions of all the models stated above to predict the outcome of a game. Post midterm we will analyze whether any of the models are overfitting or underfitting and try to improve the generalization error rate.