Research Paper Proposal

Your Name

Date

Predicting NBA games using machine learning methods

As the premier men's professional basketball league in the world, one of the four major North American professional sports leagues, NBA has huge influence around the world. NBA has 30 teams in total, 15 in each conference (eastern conference and western conference), 5 in each division (Atlantic, Central, and Southeast in Eastern conference while Northwest, Pacific and Southwest in western conference). NBA players are best paid sprotsmen and numerous of fans and experts want to analyze the outcomes of NBA games, for fun or for business, like get a better understanding of their favored teams and their opponents. So, a good prediction of NBA games outcomes would help to assess the team's performance and assist coach's decision making, or even for betting lines.

Background

Early data mining about sports are totally manually, it is time consuming and not worthworthy. With the advance of computer technology, the data mining about sports became popular. For example, people already made pretty good data mining in different sports leagues, like MFL, NCAA, and NBA.

Many studies have worked on predicting basketball game outcomes, both in NBA and NCAA. Most experts could reach an accuracy of 65%, so the models developed in different studies always want to reach at least 65% of accuracy. However, the algorithms and vectors used and the results differed. Shi Zifan's study compared Naive Bayes, Logistic regression/ Markov chain and multilayer perception to predict the NCAA match outcomes, Bryan Cheng's study used Linear regression and support vector machine to predict NBA betting lines. Loeffelholz used neural network to predict NBA games. Researchers also tried to use related attributes to build better model. For now, attributes used include team winning rate, "Four Factors", and some other basic statistics, like score points, turnover, blocks, shots made and so on.

In summary, simpliest machine learning methods are preferred because they are quite simple and the prediction accuracy is satisfactory, such as Linear Regression and Naive Bayes. In addition, SVM (Support Vector Machine) and LRMC (Logistic Regression/ Markov Chain) models are also mentioned a lot of studies. However, Neural Network is currently common known the most complex and most accurate method compared to others. Another interesting summary is many researches believe that attributed chose in the model also had an inflence on the result of

predicting accuracy and further studies could fosuc on more specialized attributes which are missing right now.

Data sources

Detailed and even some advanced specialized data could be found from the following websites:

- Basketall Reference basketballreference.com
- NBA ESPN nba espn
- HOOPS STATS hoopstats
- www.databasebasketball.com databasebasketball.com

Algorithms source

In my project, I am planning to test above algorithms to see their performance based on my data. The algorithms involved would include: Linear Regression, Naive Bayes, SVM (Support Vector Machine, LRMC (Logistic Regression/ Markov Chain) and Neural Newtork.

Next, I tried to add some attributes into the models to see if the attributes like age, injury, schedule power, experience could help to improve the model performance.

If the added attributes could improve the accuracy of the model, then I would use take-one-out approach to evaluate the importance of added attributes and possibly made a ranking for them.

The code sources for the algorithms I chose are can be found from these websites:

Linear regression

R Tutorial - Simple Linear Regression Linear regression in R

Logistic regression/ Markov chain

R Data Analysis Example: Logistic Regression Logistic regression in R

Support Vector Machine

Support Vector Regression with R VSM in R

SVM{e1071}SVM Inside R

Naive Bayes

Naive Bayes{e1071}Naive Bayes Inside R

Multilayer Perception/ Neural network

Neural Network exampleR code example

Neural Network with RA simple example of Neural Network

References

Shi, Zifan et al. "Predicting NCAAB outcomes using ML techniques - some results and lessons learned." 14 Oct, 2013. arxiv. Org

R. A. Torres. "Prediction of NBA games on Machine Learning Methdos". December, 2013. University of Wisconsin Madison. (http://homepages.cae.wisc.edu/~ece539/fall13/project/AmorimTorres_rpt.pdf)

Bryan Cheng, Kevin Dade et al. "Predicting the Betting Line in NBA Games". Stanford Univeristy (http://cs229.stanford.edu/proj2013/ChengDadeLipmanMills-PredictingTheBettingLineInNBAGames.pdf)

Matthew, Hongfei Wang. NBA Oracle (http://www.mbeckler.org/coursework/2008-2009/10701_report.pdf)

"Predicting outcomes of NBA playoffs using algorithms". (http://www.cse.usf.edu/~xqian/courses/CIS6930_prml/reports/report_mo.pdf)

B. Markoski et al. "Using Neural Networks in Preparing and Analysis of Basketball Scouting". (http://dx.doi.org/10.5772/48178)

Loeffelholz, Bearnard;Bdenar Ear;and Bauer.Kenneth W.(2009) "Predicting NBA Games Using Neural Networks," Journal of Quantitative Analysis in Sports: Vol.5 : Iss. 1. Article 7.

S. P. Kbam and J. S. Sokol. "A logistic regression/Markov chain model for nacc basketball." Naval Research Logistics, 53:788-803,2006.