Prediction of winner in NBA basketball games

Zejian Zhan

1 Motivation

The main objective of this project is to predict the winner of a NBA basketball game. The core steps are collecting data, building up models, setting up the features, designing learning algorithms and analyze the accuracy rate.

2 Data collection

Basically there are a lot of websites collecting the scores of the game, along with more detailed information like the length of a game, the recent game results of a team and so on. So the idea is to grasp the data from the website by program, for example, using python script to get data from http://www.basketball-reference.com or http://www.nbaminer.com.

3 Models

After collecting data, what is required to do is analyze it and try to visualize it. Some illustrations like histograms and scatter plots are preferred for better observation.

4 Features

To better predict the winner of a game, there are many factors worth attention. One important figure is the percentage of winning versus another team and the points difference. The percentage suggests the overall probability of which team is stronger while the points difference indicates the degree of being stronger of the winner. Besides, the players in both teams are also pretty important. Some professional players may suffer from illness so that they cannot participate into the game and this situation will weaken the probability of being the winner. Also the location of the game is also very crucial. It may be easier for people to win the game in their own city. So here it may be somewhat complicated to define the strength for players as players play different roles in a game. And it is also hard to define the ability point of a player.

5 Learning algorithms

Right now the basic idea is to set up the weight for each feature so that a mathematic formula will be obtained for the classification.

6 Evaluation

The collected data can be divided into two parts. One part is training data and another one is test data. So with the training data the parameters will be set and with the test data the prediction will be generated to compare with the actual result to calculate the accuracy.

7 Schedule

Week 3-4	Collect data from the website	
Week 5	Analyze data for data pre-processing	
Week 6	Organize dataset	
Week 7-9	Implement the core part like algorithm, train, test process and so on	
Week 10	Interpret the results	
Week 11	Improve the results by adjusting the algorithms	
Week 12	Write the final project report and prepare for the presentation	