Case Study of Prediction in Sports Games with Bigdata and Mash-Up

20201635 전찬

As Bigdata and Computational power grow rapidly, we can predict many things by using those techniques. Especially, in baseball, which is one of the most popular in Korea, it is important to evaluate a player by their past game data, or predict who will win a match between two teams. In MLB, they use ZiPS to estimate a player, and interest in prediction of a baseball game is increasing.

In this study, we estimate S.Y. Lee’s data who is the best hitter in Korea. Because it is efficient to use classification technique to predict some outcome, we use Decision Tree to which is easy to apply, and easy to use. And we use Data Mash-Up technique to integrate some data from other sources. After Web has been invented, it is easy to gather, use data in web pages. We use KBO, Korea Meteorological Administration, and Google Trends data. There are also many researches that tried to predict Baseball, Basketball result by using Empirical Bayes, ANN, SVM, Naïve Bayes. But these kind of algorithms isn’t convenient to apply Mash-Up data, thus Decision Tree can be the best way to estimate S.Y. Lee’s data.

In Implementation step, 10 decision trees are produced And 3 outcomes of them is the most important things, which is H(hit), BB(base on balls), and result. H is divided by 0, 1, 2, 3 and BB is divided by 0, 1, 2 and result is divided by 1(lose), 2(win), 3(a tie). In Conclusion, we can implement which condition is the best condition(external condition) the player can do best in game. By this information, we can predict a player’s performance. And also, by Decision Tree, we can do Data Mining that which information effects a player’s performance. And by this study, we can learn that Data Mash-Up can be used to predict particular area.