CS584 Final Project Proposal

Ting Jiang A20386491,Chen Gong A20379606, Yizhi Hong A20386348

1. Introduction

Nowadays soccer is the most popular sport game in the world. Over half of the population around the world are attracted to a soccer game. Also, 2018 is a World Cup year. More importantly, there are a huge number of statistical issues related to soccer players and teams to discover. Hence, it would be exciting and it is necessary to develop a machine learning model to show the statistical explanations and make predictions to a soccer player or a soccer team. Thanks to the video game, FIFA 18, which generously provided a dataset (fifa-18-demo-player-dataset) for this research.

2. Problem

1. The overall statistics of FIFA 18.
    a. Top 20 players in the game.
    b. The most powerful team in the game.
    c. The most powerful country in the game.
    d. The distribution of wage and value of the player.
2. How the age influencing the overall rating of the player.
3. How the age influencing the player's wage and the player's value.
4. Predict the position of the player. (Classification)
5. Predict the overall rating of the player. (Linear Regression or KNN)

3. Machine Learning Technique

Generally, the methods of regression and classification would be used in this research. To be specific, in the part of regression, we would use linear regression and k-nearest neighbors, while

in the part of classification, we would use naive bayes classifier. We would apply 3 standard-deviation to mark outlier, such as Messi, and remove them from the input data. To evaluate our results, we would apply partials from the real dataset to be test data in the developed models and compare to the real data to analyze the accuracy of our result.

4. Input

There are 17980 players in the dataset, and we will use part of the players' data as training data to fit the models. In order to proof our prediction model, we will randomly pick the data from the rest of the dataset to test our models. And we will randomly generate several virtual players to make some predictions.

5. Expected Result

For the first problem, it will output the statistical result.

For the second and third problem, we expect that age will influence the overall, wage and value.

For the fourth problem, we will use the characteristics of a player, such as the agility, dribbling, to predict the best position for them.

For the last problem, we expect that we could predict the overall rating of a given player with considerable accuracy.

References

[1] FIFA18 Dataset. https://www.kaggle.com/thec03u5/fifa-18-demo-player-dataset/data
[2] Taking a look at the FIFA 18 Player Dataset.https://www.kaggle.com/donyoe/taking-a-look-at-the-fifa-18-player-dataset
[3] Machine Learning Proposal Template. http://www.inf.unibz.it/~zini/ML/slides/project_proposal_example_2.pdf