Predicting NBA Over/Unders Using a LSTM Model

William Allstetter, Victor Chang, Justin Chu

Introduction

Due to a shared interest in NBA basketball, our group chose to implement the research paper: “The Bank is Open: AI in Sports Gambling.” The paper’s objective is to predict whether any given NBA game’s total score will be over or under the Las Vegas betting lines. Using this paper as motivation, our group attempted to predict the same.

Methodology

In order to generate NBA over/under predictions, our group implemented a LSTM neural network with two dense layers. Our model was trained on data from 5076 games from the 2016-2019 and tested against 1437 games from 2020-2021. An LSTM is ideal for models like this, where previous games can impact the results of the future ones. We grouped the games into timesteps of three before putting them into the LSTM. The LSTM output was then fed through two dense layers. In the course of our experimentation, we found a smaller learning rate (0.0001) to give the best results. Game data included specific team statistics such as points scored, points allowed, field goal percentage, free throw percentage, three-point percentage, assists, and rebounds. The game data for each team is averaged over all of that team’s games played for the season. After training, our LSTM neural network is able to predict whether the total number of points that will be scored between two NBA basketball teams is over or under the given Las Vegas Over/Under Betting Line.

Results

Loss was extremely variable across trials, but converged to a stable value as it trained. Our accuracy fluctuated from 48% to 53%, most often landing around 52%. When compared to the research paper we implemented—whose accuracy averaged 51.5%—our implementation was successful. The difficulty surrounding training a neural network to predict over/unders is that the network is competing against an intelligent adversary whose goal is to pick a line that even the most informed actors have a 50% chance of beating. Clearly, the line-pickers are good at their job, but our model was able to exploit a 2% weakness.

Challenges

One of the largest problems that we faced was obtaining the data necessary to train our LSTM neural network. We initially attempted to use a web crawler to scrape the data from the NBA statistics website Basketball Reference. However, this proved to be extremely challenging. Luckily, we were to find online the data that we needed in the form of an excel spreadsheet. Another problem that we faced was preprocessing. Our group ran into a few bugs when formatting and preparing the data for the LSTM model; additionally, it was challenging to match each set of game data with the accompanying training label (which was dictated by the Las

Vegas sports betting lines). Finally, our data set contained a lot of irrelevant game data; we solved these challenges by removing games such as preseason games and special scrimmage games during the 2020 NBA bubble.

Reflection

Overall, our group was satisfied with the progress of our project. We were happy to meet our base goal of creating a functioning LSTM neural network that generates predictions about the total number of points scored in a game. Our model worked exactly the way we wanted it to. Utilizing an LSTM neural network allowed us to account for each previous game sequentially. Additionally, we were extremely satisfied to meet our reach goal of creating a model that performed better than the model in the research paper. Our approach stayed consistent throughout the duration of the project. However, the challenges with the web crawler caused our group to pivot into a different method of obtaining the requisite game data. If we had more time and access to more data (e.g. advanced analytics data, pace, offensive/defensive rating), we feel that we could create a model that generates even better predictions. Overall, our group learned a lot from this project. It was really interesting to apply the concepts that we had learned in class to our group's shared passion.