
Optimizing Preseason Fantasy Football Rankings

John O'Hollaren (jpo4)

Department of Electrical and Computer Engineering

Abstract

Fantasy football is a fast growing, multi-billion dollar industry centered around projecting which players will perform the best over the course of a season. Correctly predicting which players will play well greatly improves your team's year-long performance, whereas wasting a high pick on a low achiever can ruin your season. This paper focuses on an optimal strategy for predicting which NFL wide receivers will have the best fantasy football season using data available prior to the beginning of the season.

1 Introduction & Related Work

According to the Fantasy Sports Trade Association, over 10 percent of the US population spent 12 billion dollars playing fantasy football in 2012. Given such a large market, a slight statistical advantage could provide significant profits in fantasy football leagues. Specifically, a correct pre-season ranking of wide receivers (WRs) will allow owners to pick the best players to join their teams during the pre season fantasy football draft, setting them up for year-long success. However, ranking players is difficult. Preseason rankings from teams of experts at ESPN and Yahoo are correct for only a small minority of players. This paper compares several machine learning techniques for ranking WRs, including linear regression, feature reduction, k-means clustering, mixture models, and PCA. A PCA regression technique which outperforms ESPN and Yahoo experts is presented.

While countless fantasy football websites exist, the major players are ESPN, Yahoo, the NFL, 4for4, and FantasyPros. None of these entities releases their methods publicly. Tim Mathews at the University of Southampton¹ explored using Bayesian reinforcement learning for week-to-week team management during the season. Matt Bookman at Stanford used linear regression to predict week-to-week performances during the season with limited success only in certain weeks². The week-to-week prediction problem, however, is simplified through the use in-season data from each player, whereas in the pre-season, year-long ranking problem, predictions must be made for players who are 8 months removed from public games. Some attempts at machine learning approaches for the year-long ranking problem have been made³, but did not outperform human experts. Other attempts use machine learning to outsmart other players in the same league⁴.

1.1 How Does Fantasy Football Work?

Points are awarded when a player scores a touchdown or gains yards. Each owner picks their team at the beginning of the year. Correctly predicting which players will play well greatly improves your team's year-long performance. This paper focuses on an optimal strategy for predicting a list of the top 30 non-rookie NFL wide receivers for a season, using only data available prior to the start of the season..

1.2 Data Summarization

My dataset includes NFL statistics for over 700 players for every year from 2007 to 2012. There are 13 features in total: Receiving Yards, Receiving Touchdowns, Receiving Targets, Receiving Catches, ESPN Preseason Ranking, Yahoo Preseason Ranking, $\Delta 1$ year Receiving Yards: (Yards in year Y-1) - (Yards in year Y-2), $\Delta 1$ year Receiving TDs: (TDs in year Y-1) - (TDs in year Y-2), $\Delta 2$ year Receiving Yards: (Yards in year Y-1)-(Yards in year Y-3), $\Delta 2$ year Receiving TDs: (TDs in year

¹Mathews, Tim, et. al. Competing with Humans at Fantasy Football: Team Formation in Large Partially-Observable Domain. Association for the Advancement of Artificial Intelligence, 2012.

²Bookman, Matt. "Predicting Fantasy Football - Truth in Data." Stanford CS 229 Final Project. December 14, 2012.

³Kapania, Nitin. "Predicting Fantasy Football Performance with Machine Learning Techniques." Stanford CS 229 Final Project. December 14, 2012

⁴Becker, Adrian, et. al. "An Analytical Approach for Fantasy Football Draft and Lineup Management." MIT ORC Technical Report. November 2010.

Y-1) - (TDs in year Y-3), $\Delta 3$ year Receiving Yards: (Yards in year Y-1) - (Yards in year Y-4), $\Delta 3$ year Receiving TDs: (TDs in year Y-1) - (TDs in year Y-4), Fantasy Points in Previous Year: $6 \cdot \text{TDs} + 0.1 \cdot \text{Yards}$

1.3 Performance Criteria and Baseline

Discounted cumulative gain (DCG) will be used to assess WR rankings. It rewards ranking good players highly and penalizes ranking bad players too high. A higher DCG is better:

$$DCG = rel_1 + \sum_{i=2}^p \frac{rel_i}{\log_2(i)}, \text{ and the relevance weights are: } rel_i = \frac{1}{\text{actual end of year rank}}$$

Using DCG, the dataset can be analyzed by looking at the performance of ESPN, Yahoo, and simple regression across the years.

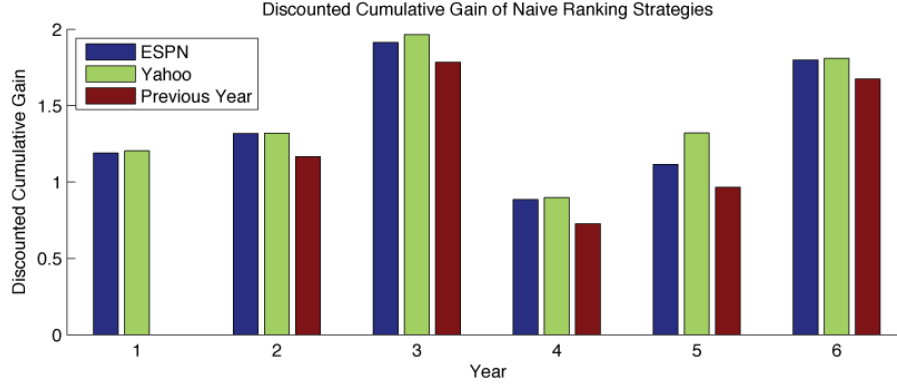


Figure 1: : DCG of ESPN / Yahoo preseason rankings over the years, along with a naive method of using the previous years final rankings as the next years preseason rankings.

2 Models and Methods

2.1 Regression

3 Results

4 Conclusions

5 Bibliography