REU 2009
Final Report

Who's #1? The Science Of Building Ranking Systems

Student Participants

— John Holodnak (Ohio Northern)

— Chandler May (Harvey Mudd)

— Daniel Moeller (Notre Dame)

Advisors

— Carl Meyer (NC State Faculty Advisor),

— Ralph Abbey (NC State Graduate Student Advisor)

Sports ranking models were the main area of research for our REU project. We studied several different models and the theory behind them. Utilizing MATLAB as a programming language, we used these models to predict game outcomes and point spreads, and compare the relative value of statistics as indicators of team strength.

Keener's Direct Method (1993) ranking model was the starting point in our research. In order to fully understand Keener's model we studied the Perron-Frobenius Theorem, which provides the theoretical basis for Keener's ranking system. We then examined Anjela Govan's Markov model (2008) and Offense-Defense model (2008). The Markov model is based on the theory of Markov chains and the Offense-Defense model is an application of the Sinkhorn-Knopp Theorem (1967). Finally, we implemented the Massey model (1997), which relies on the theory of least squares and is the basis for one of the computer models used in the BCS.

After implementing each of the models, we used NFL data from the 2002-2008 seasons to make game predictions. We compared their efficacy in terms of hindsight and foresight prediction accuracy. In hindsight prediction, we used the ranking generated from an entire season of data to predict all the games from that season. Foresight predication refers to predicting a week of games using the ranking from previous weeks. The models performed similarly in both hindsight and foresight game prediction.

Next, we attempted to use the models to predict the point spread of games. We plotted point spreads against the rating differences in the models. Lacking a clear trend or any further ideas, however, we decided that our time would be better spent on other questions. One of these questions was whether we could determine the maximum possible hindsight accuracy for a season. In doing so, we realized that higher hindsight accuracy did not necessarily correspond to higher foresight accuracy, which we consider a better measure of a model's strength. We were able to generate hindsight accuracies significantly higher than those from the previously mentioned models, but abandoned this idea due to disappointing foresight results.

The area on which we spent the most time and made the most progress was our attempt to determine whether rushing or passing yards is a superior game predictor and indicator of team strength. To explore this problem, we first applied each of the four models to rushing yards and passing yards. We also ran the Markov model with combinations of scores with rushing yards and scores with passing yards. Each application produced a higher foresight accuracy using rushing yards than passing yards. To strengthen our results, we also explored the correlation of differences in game

scores and differences in rushing and passing yards in those games. The correlation with difference in rushing yards was stronger than with difference in passing yards, which led us to believe that out-rushing an opponent is more likely to correspond with outscoring an opponent than is out-passing them. Additionally, we looked at the fractions of games a team won given that they outperformed in rushing, passing, or both. Outperforming in rushing yards resulted in a higher fraction of games won than outperforming in passing yards. Thus, all of our evidence pointed toward the conclusion that rushing yards is a superior indicator of team strength when compared with passing yards.

Our final objective was to find an optimal combination of statistics to use for the Markov model. In collaboration with the subset selection group, we performed subset selection in an attempt to pick a set of statistics that best represented all those that we had collected. It seems that a set of statistics without scores will not do very well and that any set of statistics that has scores will do fairly well.

This research is reported in the paper *Rush vs Pass: Modeling the NFL* that is being prepared for publication in the *Journal of Quantitative Analysis in Sports.*

Conclusion and Accomplishments

- We analyzed several different sports ranking models, applying them to the NFL. We then used three different approaches to examine whether rushing yards or passing yards is a better indicator of team strength. Each approach provided evidence that rushing yards is a better indicator of team strength than passing yards. The specific accomplishments are as follows.
  - — We learned about four different ranking models and the theory behind them.
  - — We compared the four models and discovered that they are competitive with each other and with ESPN analysts (Chris Mortensen and Mike Golic).
  - — We examined the correlation between score differences and rating differences and discovered that it was not strong enough to provide any significant conclusions.
  - — We worked on optimizing hindsight accuracy and concluded that optimized hindsight accuracy did not imply optimized foresight accuracy.
  - – We compared rushing yards and passing yards as measures of strength
    - ▷ We looked at which performed better in the four models
    - ▷ We determined that differences in rushing yards more closely correlates with differences in scores than passing yards
    - ▷ We saw that teams that out-rushed their opponents were more likely win than teams that out-passed their opponents
  - — We collaborated with the subset selection group to determine which subset our statistics best represents the set of statistics.
  - — The subsets they selected did not prove to be as good as the combinations using scores.

Publications

- A paper *Rush vs Pass: Modeling the NFL* is being submitted for publication in the *Journal of Quantitative Analysis in Sports.* A first draft can be obtained at `http://meyer.math.ncsu.edu/Meyer/REU/REU2009/REU2009.html`