NFL & NCAA Football Prediction using Artificial Neural Networks

Andrew D. Blaikie College of Wooster 1189 Beall Ave. Wooster, OH, 44691 ablaikie13@wooster.edu

John A. David Virginia Military Institute 512 Letcher Ave. Lexington, VA 24450 davidja@vmi.edu

ABSTRACT

The modeling of predicting the outcome of football games is very challenging. Even the best predictive models have an average absolute error of 10-12 points per game in the NFL and 12-14 points in Bowl Championship college football. Artificial neural networks were used to create models to predict the outcome of football games for both the NFL and college football. The NFL model was a continuation of the model in [1] and the college football model was new. Data analysis was done to identify the most predictive statistics, which were later used in the model. The model used was based purely on statistics and used a committee of machines approach for greater consistency. Many models were compared to determine which was the most accurate. It was found that the college football model performed poorly when compared to the NFL model. We discuss reasons for these results and procedure to overcome the challenges. Afterwards, the models were examined using derivative analysis. The results of the research showed that the NFL model consistently was in the top half compared to other prediction experts, while the college football model tended to be closer to the middle of these rankings.

Proceedings of the 2011 Midstates Conference on Undergraduate Research in Computer Science and Mathematics Gabriel J. Abud College of Wooster 1189 Beall Ave. Wooster, OH, 44691 gabud14@wooster.edu

R. Drew Pasteur College of Wooster 1189 Beall Ave. Wooster, OH, 44691 rpasteur@wooster.edu

Key words: Artificial Neural Networks, NFL, NCAA Football, Prediction

1. INTRODUCTION

Football has been a very popular sport in the country since its introduction. The Super Bowl is consistently one of the most watched television events in the U.S. and both the NFL and college football have millions of fans. Predicting the outcome of football games has never been straightforward. There is a high level of randomness in football games that makes it difficult to obtain consistently accurate predictions. In addition, some statistics such as injuries and starting lineups are hard to quantify in a way that can be put into a mathematical model. The best predictive models as seen on *thePredictionTracker.com* [2] have an average absolute error of 10-12 points per game in the NFL and 12-14 points in college football.

ANNs are applicable to many areas of science. Some of the practical applications of ANNs include vehicle control, quantum chemistry, game playing and decision making, face and speech recognition, medical diagnosis, and e-mail spam filtering [3]. They are especially good when trying to find a connection between a set of inputs to a set of outputs that may not have a well defined theory. We used artificial neural networks (ANNs) in our model because they have shown the ability to predict complicated systems in many applications.

We found that the model that performed well for the NFL did not have as accurate of predictions in the college football model. College football turned out to be much more difficult to predict. The difference in the accuracy of predictions is also reflected in other top models. As previously mentioned, the top models for the NFL were on average 2 points more accurate than the top college football models. There are several reasons for this decreased accuracy. Among these are a much wider range in the abilities of teams and differences in each team's strength of schedule.

In this paper we compare different types of ANN models on both NFL and college football games. The paper is organized as follows. Section 2 contains a description of previous work in ANN prediction involving sports. We then describe the data collection process in Section 3. In Section 4 we describe breifly the structure of Artificial Neural Networks. The different types of models tested are described in Section 5. Next in Section 6 we describe the results of the different model types for NFL prediction. Following in Section 7 we describe the results and difficulties of NCAA prediction. In Section 8 we describe the model and data statistical analysis implemented.

2. PREVIOUS RESEARCH

Artificial neural networks have been studied for decades and have a significant body of literature. Even so, it has proven difficult to apply them to real world applications. The research by Purucker [4] introduced the idea of using ANNs to predict the outcome of NFL games. This research offers an excellent introduction to using ANNs in football predictions. Purucker applied several types of supervised and unsupervised networks to predicting weeks 15 and 16 of the 1994 NFL season. Work by Kahn [5] used 13 weeks of game data, including total yardage differential, rushing yardage differential, time of possession differential, turnover differential and home field advantage to predict weeks 14 and 15. He achieved 62.5% and 75% accuracy in these two weeks. Considering the ESPN experts in 2009 had between 61% and 67% accuracy predictions, this methodology shows promise. Loeffelholz et. al. performed a more extensive study of neural network prediction of NBA games in [6]. The work in this paper incorporated ideas of these past studies with several key additions including seasonto-date statistics combined with previous seasons, a simple method for incorporating team statistics, a robust method of combining multiple ANNs and analysis of the performance of this methodology

over several seasons.

This paper is an expansion on previous research from David *et. al.* [1]. They were able to achieve a high accuracy in the NFL using a committee of committees technique with artificial neural networks. We expanded on this model to include college football and also refined the existing model.

3. STATISTICAL DATA

The statistics we gathered were mostly from [7]. We collected a total of 46 statistics per game from this site for the NFL, and 40 for college football. We also gathered average stadium attendance, the previous season rating, and the opponent's win percentage for both home and away teams of each college football game for a total of 46 statistics. The previous season's ratings in the Sagarin Ratings were gathered from [8]. We collected this game-by-game data for every season from 2003 to 2010 for both the NFL and college football.

All of this was done with Perl scripts and then written to Excel files. Simple Perl modules were installed from [9] (LWP and SpreadSheet) to work with online sources and to write to Excel files. After retrieving the raw data, we then calculated the season-to-date averages for each team. This was done using both Perl and MATLAB scripts. The results were separated by week and by the season.

The last step before using the data as input to the ANN was to average each of the expected offensive statistics with the expected defensive statistics of the opponent. So for example, if team A and team B are playing each other, instead of just using the number of points team A is expected to score (their current season-to-date average for scoring) as input, we would average the number of points team A is expected to score with the number of points team B is expected to allow. If team A is expected to score 20 points and team B is expected to allow 10 points, we would average the two and say that team A will score 15 points that game. This final number was the one that was used as input to our model. These calculations were all done with MATLAB scripts.

4. ARTIFICIAL NEURAL NETWORKS

Inspired by the biological neural networks of the brain, an Artificial Neural Network (ANN) is ultimately a function, where the input is a vector of input data. This network is composed of "neurons" or processing units which are highly connected. The neurons the ANN can be taught specific patterns by feeding the network input vectors, and targeted output vectors. Given enough data the ANN can be trained to approximate additional data vectors or find relationships between input and target vectors [10].

An artificial neuron can be understood as a mathematical function or object that takes input data and produces some output. Suppose that we have a *n* element vector as the input data, the neuron will scale each data element r_j , by a weight w_j . The scaled data is summed and offset by some bias *b* and then put through a differentiable transfer function *f*. The neuron output can be viewed analytically as

$$f(r_1w_1 + r_2w_2 + \dots + r_nw_n + b). \tag{1}$$

The neuron is able to interpret the data differently depending on the weights and bias. A stronger weight for a given data point allows that data point to change the output more dramatically.

The transfer function f can be linear f(x) = x, or something nonlinear. Nonlinear transfer functions can be useful for understanding nonlinear relationships. For the models used we implemented "Tan-Sigmoid" transfer functions,

$$f(x) = \frac{2}{1 + e^{-2n}} - 1,$$
(2)

which map the input onto the interval (-1, 1).

Now that we have introduced the basic building block of our network, we construct the feed forward three layer ANNs used in the more successful models. The first layer is constructed of 4 "hidden" neurons. In the hidden layer we implement the nonlinear transfer function "Tan Sigmoid". There exists a weight between every element of the input vector and every hidden neuron. The outputs of this layer is passed as the input vector into a hidden layer of 6 neurons with "Tan Sigmoid" transfer functions. The outputs of this layer are put through an output neuron with a linear transfer function and may be compared to the target vector and the squared error gives the error function. The error function is what we attempt to minimize with respect to the weight matrices. By optimizing the error function the ANN is being "trained".

Before we train the ANN we must first prepare both the network and the data for training. First the data is partitioned into 3 sets, the training set, the validation set, and the testing set. We use 70%for the training, 15% are used for the validation, and 15% for testing set. This is done randomly. The training data is used to create the error function. Each input and target vector creates an error function which is then averaged among all the other training data yielding the mean squared error. Iterative methods are used for optimization and the mean squared error of the validation set is plotted along with the training set. The ANN can fit the training data arbitrarily well but can easily over-fit the training data yielding higher error functions for the problem the ANN is trying to approximate. For this reason the training is stopped when the error function on the validation set is minimized. The testing data is used to test the ANN as it has not been involved in the training process. The weights and biases were updated to Levenberg-Marquardt optimization.

Basic ANNs still had a good deal of variability in predictive ability. For example networks with the same structure trained on different data sets may make substantially different predictions in future seasons. This is because different ANNs partition the input data into the 3 data sets randomly. In order to create an algorithm that will give more consistent and robust results, we took an approach generally referred to as a committee of machines (CoM). In this approach many networks are trained against different random partitioning of the data set. Then based on the mean square error against the testing data, the top models are chosen to use predictively. This approach was developed in [1].

To get the most robust approach to NFL game prediction, we used a committee of committees approach where many committees' predictions are combined to form our final prediction. Using this approach with 500 ANNs in the training stage, the best 100 were used in each committee, then 50 such committees were used to achieve the results below. The mean was used to form each committee vote and to combine the committees' predictions.

5. DIFFERENT MODELS

As mentioned above, we have a total of 46 statistics for the NFL and NCAA. To create an accurate model we must either reduce the number of statistics to a manageable amount or create a large enough neural network that can find patterns in the large amount of data. We tried 5 different models for the prediction of football games. These models varied mainly in terms of the data reduction techniques implemented.

The first model we tested, used in [1], utilizes only passing yards, rushing yards, points, interceptions, and fumbles. For simplicity we will refer to these statistics as *on the field statistics*. Since this paper is a continuation of [1], these statistics were a good starting point.

The next model used another set of statistics called *efficiency statistics* which include yards per pass play, yards per rush play, points, fumbles per play, and interceptions per play. These statistics are strongly correlated with winning [11]. Teams with high efficiently statistics often have a higher win percentage. For example, passing yards has a correlation coefficient of 0.31, while passing yards per attempt has a correlation coefficient of 0.61 [11].

Principal component analysis (PCA) was another method we used to reduce the data being used in the model. PCA takes the 46 stats, normalizes them and creates a new lower dimensional data set made up of linear combinations of the old statistics. The new statistics are orthogonal to each other and ordered in such a way that the first principal component accounts for the maximum variability in the data. By computing this transformation we are able to discard all but 14 principal components and keep 90% of the variance. This makes a simpler problem for the artificial neural network. This procedure also tells us that 46 independent statistics gives redundant data to differentiate games. The mathematics required was implemented using a software package in Matlab.

Using a technique that we call linear regression combinatorial optimization we came up with a set of *LRCO statistics*. This method tests every possible set of statistics to identify the optimal combination. We started by getting the statistics from the previous season and used them to train linear regression models. The models were then tested on the following season. We repeated this procedure for every season and on every possible combination of statistics. The average error for each season was used to judge which statistics were best for the model. Since linear regression is many times faster than ANNs, the computation time required to run every combination of statistics was manageable. The results are shown in Table 1.

The results were not entirely expected. No previous research shows that 4th down conversions or

Table 1: Optimal combination of statisticsin a linear regression model.

NCAA	NFL
Scoring	Scoring
4th Down Conversions	Completed Passes
4th Down Attempts	Recovered Fumbles
Net Rushing Yards	Interceptions
Turnovers	Sacks
Opponent's Win Percentage	
Rating from Previous Season	

attempts have a big impact on the outcome of a game. These may have been identified as useful by our models because of over-fitting or because they represent a new variable to the game. For example, scoring and time of possession both have a strong correlation of winning, but using both of them in a model might be redundant.

Another data set tested used every statistic gather. For this model we increased the size of the ANNs to 8 neurons in the first hidden layer and 4 neurons in the second hidden layer. This model was not found to be very successful.

6. NFL RESULTS

Fig. 1 gives a table comparing our different models to that of the other computer based NFL predictors. To test how our models perform we compare

Mean Absolute Error						
Model	2007	2008	2009	2010	Mean	Legend
On the Field	11.00	11.31	11.68	10.99	11.24	Upper Quartile
Efficiency	11.00	11.70	11.69	11.25	11.41	Top Half
PCA	11.21	11.12	11.48	11.20	11.25	Bottom
Every Statistics	11.26	11.24	11.68	11.18	11.34	Half Lower
LRCO	11.20	11.10	11.51	11.16	11.24	Quartile

Figure 1: The cells represent the mean absolute error and the color shows how the model performed in comparison to the other experts. White means our results were in the upper quartile of statistical models, light gray means the top half, etc. The mean column is not color coded according to legend. On the field statistics, PCA, and the LRCO statistics performed reasonably well. to other computer based simulations on *thepredictiontracker.com* [2]. This site lists all the top models for predicting NFL and college football games. By looking at only the models who predict every game we are able to compare how our prediction did in relation to the mean, upper, and lower quartile.

The results from the NFL show that the data reduction techniques generally work better than using a larger network and all of the data. It was also found that the LRCO statistics performed with the lowest mean error over all the seasons, but not by much.

It can be seen that a given model does not have a constant performance over the various seasons. There is variability in the outcome of football games, one model may perform very well in one season but poorly in the next. However Fig. 1 shows that our models generally perform in the top half in comparison to other models indicating that we have a generally well performing model for NFL prediction.

7. NCAA RESULTS

The results of using a top model in the NFL on the NCAA is shown in Fig. 2.

Legend Quartile						
Mean Absolute Error						
2005	2006	2007	2008	2009	2010	
13.87	12.73	13.51	14.04	12.71	13.89	

Figure 2: Table of the on the field statistics model's performances over various seasons. The cells give the mean absolute error and the color how the model performed in relation to the experts. The mean column is not color coded according to legend. This particular model performed quite poorly.

By adding additional combinations of statistics we improve the results as shown in Fig. 3. The statistics that resulted from the linear regression optimization performed with the lowest mean error. These additional statistics improve the performance of the model but it does not do as well as the NFL models. To remedy the initial poor performance of

	Legen	d Uppe Quart	Ion H	alf Botto Hal			
	Mean Absolute Error						
Model	2005	2006	2007	2008	2009	2010	Mean
Attendance	13.89	12.55	13.31	14.14	12.74	13.78	13.40
Attendance & Team Rating	13.68	12.39	12.98	14.05	12.31	13.66	13.17
Team Rating & Strength of Schedule	13.62	12.43	13.13	13.70	12.20	13.54	13.10

Figure 3: Table of the various models performances over various seasons. The cells give the mean absolute error and the color how the model performed in relation to the experts. Shows the accuracy of predictions after adding in additional statistics to improve the college football model. The mean column is not color coded according to legend. These statistics attempt to remedy strength of schedule differences and differing conferences levels.

our model we tried adding combinations of three different statistics that were not previously considered, these include the average attendance of the teams home stadium, the rating from the previous season from USA Today, and the given team's opponent's win percentage. The USA Today rankings were based on previous season statistics. These were used to give our new model an idea of the different strengths of teams from the start. For future work we would generate unique rankings instead of using the USA Today ranks to make the model purely based on statistics.

The USA Today rankings and the average attendance attempt to remedy the situation where a team from a tough conference plays a team from a weaker conference. The opponents win percentage gives a way to deal with the situation where two teams have not played even opponents giving skewed statistics. Because of the high degree of parity in the NFL these were not deemed as necessary but they could be considered.

In college football we expected that a purely statistical model would not properly describe how a team is expected to perform. For example, the NCAA has many conferences with varying degrees of team strength. When a team from a strong conference plays another team from a weaker conference a purely statistical model based on on the field statistics would predict a relatively even match. However, this is surely not the case. Another example is that in the early season a strong team could play a weaker team and rack up rushing yards, points, passing yards, etc. Then that team would have very impressive season to date statistics. When that team plays a team of the same caliber who has had more difficult opponents a statistical model would give the team who has played weaker opponents the advantage when in fact, these two teams are an even match. To confirm these ideas we tested a college football model built on the on the field statistics, Fig. 2. The results from this model confirm that a purely statistical model is not sufficient for NCAA prediction with ANNs.

8. STATISTICAL ANALYSIS

One of the techniques we used in understanding our model was a derivative-based technique. In the case of ANNs, the derivatives tell us how much a change in each input will affect the output. Essentially, this will tell us which statistics are the most important. A similar methodology is used to analyze the most important factors of a deterministic model of HIV dynamics in [12]. Figure 4 is a visual representation of how this technique works using ANNs.

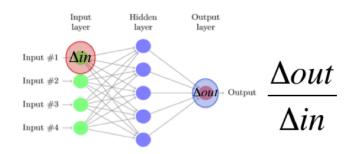


Figure 4: Visual representation of the derivative process.

Using this technique is common, however there are a few challenges using this approach on our model. Since ANNs are nonlinear functions, their derivatives will not be constant. For example, two different games will yield different derivatives. The second challenge is that our final model has 5000 ANNs. Each network will yield a different output and a different set of derivatives. To overcome this we looked at the average derivative of our model when all of the statistics are season averages. We then normalized the derivative by the input value, giving the partial derivatives the same units.

To calculate, the partial derivative was calculated for each ANN and then averaged. This was done for each statistic. Once the statistics were normalized we were able to see what percentage each statistic makes up the derivative information or the sum of the normalized derivatives.

We examine the derivatives of the model using the LRCO statistics as it performed the best in both college football and the NFL. Table 2 and Table 3 gives a chart of this info. In both models scoring makes up a very large part of the derivative information. The derivatives also showed that the rating from the previous season is very significant in NCAA games.

Table 2: Percentage of derivative information for college football. The higher the percentage the more a change in that input will affect the output of the model. The statistics shown to be significant include scoring and rating from the previous season.

Statistic	Percentage
Home scoring	24.64
Home 4th down conversions	2.77
Home 4th down attempts	3.59
Home rushing yards	1.41
Home turnovers	4.6
Home opponents win percentage	3.14
Home rating from previous season	15.52
Away scoring	17.61
Away 4th down conversions	1.10
Away 4th down attempts	0.58
Away rushing yards	3.23
Away turnovers	2.36
Away opponents win percentage	3.82
Away rating from previous season	15.58

Due the large amount of data collected over the various seasons of the NFL and NCAA, it seemed appropriate to run some statistical data tests on the large data set. The first is to calculate the correlation coefficients of a given statistic with the point differential of the game. The statistics used in this case are not the season to date averages used for prediction but are the statistics generated Table 3: Percentage of derivative information for the NFL. The higher the percentage the more a change in that input will affect the output of the model. More than 66% of the derivative information comes from scoring.

Statistic	Percentage
Home scoring	32.58
Home completed passes	5.16
Home recovered fumbles	0.61
Home interceptions	1.64
Home sacks	3.95
Away scoring	34.07
Away completed passes	9.20
Away recovered fumbles	4.27
Away interceptions	5.46
Away sacks	3.06

during a game compared with the outcome of the game. We correlate each home and away statistic with both a win/loss binary variable and the point differential. In addition we find the margin for each statistic between home and away and correlate these with the win/loss variable and with the point differential. For example correlating the scoring margin with the point differential will yield a 1 because in fact they are the same statistic. Table 4 shows the top coefficients from the set of statistics.

Table 4: Correlation coefficients for the NFL and NCAA. The higher the coefficient, the more correlation between the statistic and winning.

0			
NCAA			NFL
Total Yards	0.77	Yards/Pass	0.67
Yards/Pass	0.67	Rushing Attempts	0.66
Rushing Yards	0.65	Turnovers	0.62
Yards/Rush	0.61	Total Yards	0.60
First Downs	0.60	Rushing yards	0.57
Rating	0.52	Interceptions Thrown	0.55

These top coefficients are strongly correlated with the result of each game. The lower the coefficients, the less of an impact that statistic will have on the outcome. However these coefficients do not necessarily tell us what statistics will be good to use in a prediction. Correlation does not imply causation. For example, rushing attempts has a very high correlation with winning. This is usually due to the fact that teams who are already winning will run the ball more than pass. So they are running because they are winning, but they are not necessarily winning because they are running.

9. SUMMARY AND FUTURE WORK

In conclusion, the NFL model was improved slightly from the one created last year. The college football is not as accurate as the NFL model and and modeling college football appears to be substantially more challenging problem. The data analysis we did allowed us to decide which statistics are the most predictive and resulted in a slight improvement of the existing models. The results indicated that ANNs are a reasonable approach to football prediction. Future work will include attempting to improve the NCAA football model, and also building new models for other sports, including NBA and NCAA basketball, and Major League Baseball.

10. REFERENCES

- David, John A.; Pasteur, R. Drew; Ahmad, M. Saif; and Janning, Michael C. (2011)
 "NFL Prediction using Committees of Artificial Neural Networks," Journal of Quantitative Analysis in Sports: Vol. 7: Iss. 2, Article 9.
- ThePredictionTracker, (2007), "Football Predictions", World Wide Web electronic publication,
 www.thepredictiontracker.com/.
- [3] Wikipedia, (2011), "Aritficial Neural Network", World Wide Web electronic publication, http://en.wikipedia.org/ wiki/Artificial_Neural_Network.
- [4] Purucker, M.C. (1996), "Neural Network Quarterbacking", IEEE Potentials.
- [5] Kahn, J. (2003) "Neural Network Prediction of NFL Football Games," World Wide Web electronic publication, homepages.cae.wisc. edu/~ece539/project/f03/kahn.pdf.
- [6] Loeffelholz, Bernard; Bednar, Earl; and Bauer, Kenneth W. (2009) "Predicting NBA Games Using Neural Networks", Journal of Quantitative Analysis in Sports, : Vol. 5, number 1.
- [7] ESPN, (2011), "NFL Scoreboard", World Wide Web electronic publication, www.scores.espn.go.com/nfl/scoreboard
- [8] USAtoday, 2011, "Jeff Sagarin Computer Rankings", World Wide Web electronic publication,

www.usatoday.com/sports/sagarin.htm

[9] CPAN, (2011), "Comprehensive Perl Archive

Network", World Wide Web electronic publication, www.cpan.org

- [10] MathWorks, 2011, "Neural Network Toolbox", World Wide Web electronic publication, http: //www.mathworks.com/help/toolbox/nnet/
- [11] Advanced NFL Stats, 2011, "What makes teams win?", World Wide Web electronic publication, www.advancednflstats.com/2007/07/ what-makes-teams-win-part-1.html
- [12] David, John A.; Tran, Hien T.; and Banks; H. T., (2009), "HIV Model Analysis and Estimation Implementation Under Optimal Control Based Treatment Strategies", *International Journal of Pure and Applied Mathematics:* 57,357-392.