Any Given Sunday:
Forecasting NFL Quarterback Performance

Steven Hoerning Bobak Moallemi Matthew Wilson

December 13, 2014

Abstract

We used multinomial logistic regression and support vector machines to predict the number of touchdowns a quarterback will throw in an upcoming game. Our best model correctly predicted a quarterbacks passing touchdowns 38.6% of the time with a mean square error of 1.58. This compares favorably to ESPN's forecasts, which in 2014 correctly predicted a quarterback's passing touchdowns 33% of the time with a mean squared error of 1.39.

1. Introduction

Fantasy sports is a broad genre of games in which participants act as owners, assembling teams of players to compete against other fantasy owners based on the measured individual performance of real players. Fantasy football is the most popular version of this game among Americans. Close to 33 million people participated in fantasy football leagues last year, spending around $800 million on fantasy media products. In fantasy football, participants compete head to head by selecting NFL players to fill a lineup for a given week. At the conclusion of each game, each player's performance is translated into points by a predetermined formula. For example, quarterbacks get 4 points for touchdowns and .04 points per passing yard. The fantasy player with the most points after the conclusion of a week wins the matchup.

In this project, we used machine learning to predict the number of passing touchdowns a quarterback will throw in an upcoming game. To accomplish this we used multinomial logistic regression and support vector machines. Forward selection was used to choose a set of predictive features from approximately 25 features related to a quarterback's passing touchdowns. Our models were evaluated based on their correct prediction rates and the mean squared error of the predictions. We found that the prediction accuracy of our best models compared favorably to ESPN's predictions.

2. Data and Preprocessing

Detailed NFL data is available from several data providers, but often at high cost. Fortunately, the website www.pro-football-reference.com has a relatively complete collection of NFL data available at no charge. In order to avoid violating their terms of service, which explicitly prohibit automated scraping, we manually downloaded data from the 2008 through the 2013 NFL seasons. After formatting, our data consisted of two tables - one containing team specific data and the other containing player specific data. The team data includes 52 features that correspond to what a particular team did in a particular game. For example, this data includes features such as the total rushing and passing yards a team's offense accumulated. The player data includes 65 features that represent a player's performance in a particular game. This includes variables such as touchdowns and passing yards, as well as more detailed information such as where receivers were targeted on the field. In total, our data consists of 3,200 team performances and 30,000 player performances.

3. Models and Features

We approached our classification problem with both multinomial logistic regression and support vector machine models. For both types of models, our features consisted of rolling averages of: statistics specific to the quarterback, statistics specific to the quarterback's offense, and statistics specific to the opposing defense. Statistics specific to the quarterback include passing yards, passing touchdowns, interceptions, pass completions and pass attempts. Statistics specific to the quarterback's offense include turnovers, rushing yards, total yards, sacks, and points scored. Statistics specific to the opposing defense include total yards allowed, points allowed, sacks, and passing yards allowed. We decided to use a rolling average of these relevant statistics because each of them can vary greatly from one game to the next. We evaluated our models based on the proportion of quarterback passing touchdowns correctly predicted and the mean squared error of our predictions, both of which were estimated using using k-fold cross validation (with $k \geq 25$). Features were selected using forward selection for multinomial logistic regression and backward selection for support vector machines. For both the forward and backward selection we attempted to optimize the proportion of correct predictions.

4. Results and Discussion

Our support vector machine models and multinomial logistic regression models predicted passing touchdowns with similar accuracy. Multinomial logistic regression in general predicted more touchdowns than SVM and produced predictions with smaller mean squared errors. We creating features for both models we varied the number of previous games used in rolling average calculations. In the tables below, the number of games used to create the rolling averages is represented by m.

4.1. Multinomial Logistic Regression

The results of our best logistic regression models for different values of m can be seen in the table below. Note that the best model as measured by proportion correct used only the most recent game to make its prediction ($m = 1$).

The best set of features found by forward selection can be seen in Table 1. Notice that the best feature set varies depends greatly on m. That said, it is also important to realize that many of the statistics considered for selection are highly correlated. For example, the number of total yards a defense allows is high correlated with the number of passing yards a defense allows.

As the first Table 1 shows, the best multinomial logistic regression model (as measured by proportion correct) coincided with $m = 1$ and was correct 38.6% of the time with a mean squared error of 1.59. The confusion matrix for this model can be seen below. Note that this model

Table 1

m	Proportion Correct	Mean Squared Error	Observations
1	0.3862	1.5873	3205
2	0.3765	1.5053	2847
3	0.3711	1.5392	2530
4	0.3710	1.4772	2232
5	0.3723	1.4799	1958

Table 2

m	Features		
	Quarterback (Passing)	Offense	Opposing Defense
1	Completions, Yards	Rushing Yards, Total Yards	Total Yards
2	Completions, Interceptions	Sacks, Turnovers	Passing Yards
3	Yards, Interceptions	Rush Attempts	Total, Rush Yards, Rush Att.
4	Attempts	Points Scored	Sacks, Total Yards
5	Attempts, Yards, Touchdowns	Points Scored	Rushing Yards, Points Allowed

(and all of the other logistic regression models) never predicted more than two touchdowns for a quarterback in a game. This could be considered a shortcoming of our models given that elite quarterbacks regularly pass for three touchdowns in a game.

Table 3

True TDs	Predicting TDs		
	0	1	2
0	**0.4738**	0.2322	0.1304
1	0.2853	**0.3588**	0.2608
2	0.1605	0.2395	**0.3526**
3	0.0496	0.1247	0.1787
4	0.0229	0.0371	0.0628
5	0.0050	0.0059	0.0144
Frequency	785	2213	207

4.2. Support Vector Machines

In this section, we present our predictive results using a support vector machine (SVM) with a linear kernel.[1] As was done for the logistic regression, we tested our SVM using different values of m. Below, we list the features used within each iteration of the SVM, as chosen by backwards feature selection: Just as logisitic regression, the feature selection for the SVM varies greatly

Table 4

m	Features		
	Quarterback (Passing)	Offense	Opposing Defense
1	Yards, Completions		
2	Yards, Completions, Touchdowns, Attempts	Total Yards, Points Sored	Rushing Touchdowns Allowed
3	Yards, Completions		
4	Yards, Completions, Touchdowns, Attempts	Total Yards, Points Sored, Sacks Allowed	Rushing Touchdowns Allowed
5	Yards, Completions, Touchdowns, Attempts, Interceptions	Total Yards, Points Scored, Sacks Allowed, Turnovers, Rushing Yards, Rushing Touchdowns, Rushing Attempts, Home/Away	Passing Yards, Points Allowed, Completions Allowed, Total Yards Allowed, Passing Touchdowns Allowed, Interceptions, Sacks, Passing Attempts Allowed, Rushing Attempts Allowed, Rushing Yards Allowed

with m. That said, the quaterback's passing yards and passing completetions appear as features for each m. Note that for $m = 5$, backwards selection simply chooses all the available features. The following table details the predictive success of each SVM: Proportion Correct equals the

Table 5

m	Proportion Correct	Observations
1	0.3829	2938
2	0.3658	2600
3	0.3634	2309
4	0.3592	2038
5	0.3602	1785

fraction of correct predictions made by the SVM, computed using k-Folds Cross Validation with 10 folds. Consistent with our results from the multinomial logistic regression, The SVM with $m = 1$ seems to perform the best. In fact, the proportion of correct predictions is nearly

[1] We experimented with a number of other kernels and found that the linear kernel performed best during cross validation.

identical, at 38.29%. Below, we include the confusion matrix corresponding to our SVM model with $m = 1$. The SVM does quite well when predicting a low number of passing touchdowns.

Table 6

True TDs	Predicting TDs		
	0	1	≥ 2
0	**0.5342**	0.2286	0
1	0.2750	**0.3481**	0
2	0.1261	0.2522	**0**
3	0.0315	0.1267	0
≥ 4	0.0333	0.0444	0
Frequency	571	2367	0

The model is correct more than 50% of the time when it predicts zero touchdowns. However, the model does extremely poorly for outcomes beyond one touchdown. In fact, it never predicts a quarterback throwing more than two. While the logistic regression also fails to predict a high number of touchdowns (say, three or more), it still predicts some two touchdown events, thus outperforming the SVM here.

5. Conclusions

We used multinomial logistic regression and support vector machines to predict the number of touchdowns a quarterback will throw in an upcoming week. Our best model correctly predicted a quarterbacks passing touchdowns 38.6% of the time with a mean square error of 1.58. This compares favorably to ESPN's forecasts, which in 2014 correctly predicted a quarterback's passing touchdowns 33% of the time with a mean squared error of 1.39. That said, we are not ready to conclude that our best model is superior to ESPN's predictions because we have not yet tested our models on the data from 2014.

6. Future Work

We would certainly like to further compare our best models to ESPN's model. The most direct way to do this is to gather data on 2014 and predict the outcome of these games using our models trained on data from 2008 to 2013. The predictions of our models would be directly comprable to those of ESPN. Moving forward, we also intend to apply techniques similar to those used in this project to predict other discrete fantasy football statistics such as quarterback interceptions and touchdowns by non-quarterbacks.