

Estimating the Probability of Winning an NFL Game Using Random Forests

Dale Zimmerman

October 4, 2023

Brian Burke's NFL win probability metric

- May be found at www.advancednflstats.com, but the site has been inactive since Burke joined the ESPN Analytics Department in 2015
- Bins every play in a training dataset according to current score, time remaining, and field position
- Estimates win probability (WP) for a new play by the proportion of training observations in the corresponding bin for which the team on offense won the game, with some adjustment for down and yards to go for a first down
- Some extrapolation and smoothing used to borrow strength from bins similar to the one corresponding to the new play

Lock and Nettleton (LN) random forest approach

Similar in some ways to Burke's, but differs as follows:

- Replaces subjective binning of training data with data-driven partitioning that minimizes prediction error (random forest)
- Additional variables are included: pre-game point spread, a variable that combines difference in score with time remaining, number of remaining time-outs each team has, and total points scored
- Allows for assessment of relative importance of each variable

Data

Training data:

- Data from all plays from 2001-2011 NFL seasons
- $n = 430,168$ plays
- $p = 10$ predictor variables
- Response y_i is an indicator variable for whether the team on offense before play i won the game

Test data: Data from all plays from 2012 season

Regression trees

- A classification or regression tree is “grown” by performing a series of binary splits, each of which splits the p -dimensional space of predictor variables into two parts according to the value of just one of the variables
- At each split, the single variable is selected and the split-point is chosen to minimize the misclassification error (a “greedy” algorithm)
- Stops when when a pre-set minimum node size is reached
- The end result is a set of (hyper)rectangles corresponding to the terminal nodes; some “pruning” may then be performed
- For a regression tree, the predicted response is the average of the observations in each terminal node

Two-predictor illustration (from *Elements of Statistical Learning* by Hastie, Tibshirani, and Friedman):

306 9. Additive Models, Trees, and Related Methods

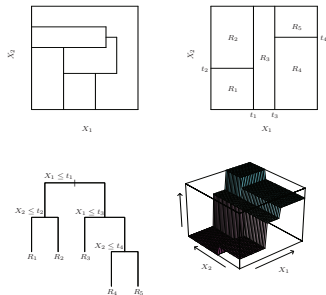


FIGURE 9.2. Partitions and CART. Top right panel shows a partition of a two-dimensional feature space by recursive binary splitting, as used in CART, applied to some fake data. Top left panel shows a general partition that cannot be obtained from recursive binary splitting. Bottom left panel shows the tree corresponding to the partition in the top right panel, and a perspective plot of the prediction surface appears in the bottom right panel.

Bagging and random forests

- Predictions corresponding to an input \mathbf{x} obtained by trees generally have low bias but high variance.
- To reduce this variance, *bagging* (bootstrap aggregation) is a useful technique: we fit the same regression tree many times to bootstrap samples of the training data, and average the predictions.
- This variance may be reduced even more if we can reduce the correlation between the trees.
- Random forests achieve this by, before each split, randomly selecting a subset of the predictor variables as candidates for splitting.

Steps of LN random forest approach

1. Draw a bootstrap sample (of size 430,168) from the training dataset and regard these data as belonging to a single node N_0 .
2. Randomly select m (taken to be 2) predictor variables from the 10 predictors.
3. For each selected predictor variable x and all possible “cut-points” c , compute the sum of squared errors.

$$\sum_{k=1}^2 \sum_{i \in N_k} (y_i - \bar{y}_k)^2,$$

where N_1 and N_2 are the subsets of training observations with $x \leq c$ and $x > c$, respectively, and \bar{y}_k is the average of the y_i 's in N_k .

Steps of LN random forest approach, continued

4. Choose the variable and the cutpoint c to minimize the sum of squared errors in Step 3, and split the training data into two subnodes accordingly.
5. Repeat Steps 2 through 4 recursively at each resulting subnode until either:
 - (a) the number of observations in any subnode is less than *nodesize* (a user-supplied tuning parameter taken to be 200), or
 - (b) the y_i 's corresponding to all the observations in a subnode are identical.

These 5 steps produce a tree. Repeat them B times to obtain a random forest.

Prediction using the random forest

To predict WP using the random forest:

1. For each tree, trace the path of the new play situation down the branches of the tree to a terminal node, and obtain the predicted response for that tree as the proportion of wins in that terminal node.
2. Average the predicted responses across all B trees in the forest.

Assessing relative importance of predictor variables

Methods:

1. Graphical assessment

- Plot how WP changes when one variable is changed while holding all the others constant
- See Figure 3 of paper

Assessing relative importance of predictor variables

2. Numerical assessment:

- Randomly permute the values of predictor variable k with the test set and re-predict WP;
- For each play i compute the squared error after permutation to the original square error;
- Repeat the above two steps 100 times and compute the average increase in squared error for play i ;
- Average over all plays to obtain a measure of overall importance of the k th predictor variable.
- Overall, difference in score and spread were the two most important variables.

Addressing other interesting questions

- Can chart the WP before each play of a game to see how it evolves over the course of the game (i.e. a time series plot)
- Can determine which play over the course of the game, a season, etc. resulted in the greatest change in WP
- Can use change in WP after each of several successful play options to assess coaching decisions. For example, in the 2016 AFC Championship Game, Broncos vs. Patriots, the Patriots were favored by 3. With 6:03 left in the game and 3 timeouts remaining, they faced 4th and 1 from the Broncos 16 trailing 20 to 12. Should they go for it or kick a field goal?

Punt, kick a field goal, or go for it?

If we know, for each of these 3 options,

- the occurrence probability of each potential outcome (punt returned for touchdown, or punt and net yardage; field goal made or not made; touchdown scored, first down made or not made plus yardage, or touchdown scored against); and
- the offense's WP for each new game situation that results from each outcome,

then we may compute the expected win probability (EWP) corresponding to that option as

$$\text{EWP} = \sum (\text{Probability of outcome}) \times (\text{Corresponding WP}).$$

Punt, kick a field goal, or go for it?

For example, EWP for a field goal at a given distance is

$$\text{EWP}_{FG} = \theta \cdot \text{WP}_{make} + (1 - \theta)\text{WP}_{miss}$$

where θ = probability of a made field goal (at the given distance).

We may use outcomes of all previous field goal attempts (possibly smoothed by a logistic regression of those outcomes on distance) to estimate θ at the given distance.

This approach is due to Zimmerman and Nettleton (2016, Midwest Sports Analytics Meeting).

Back to the 2016 AFC Championship game...

- Coach Bill Belichick elected to go for it, and the Patriots were stopped short. However, it was the right call at the time since $EWP(\text{field goal attempt}) = 0.199$, while $EWP(\text{go for it}) = 0.281$.
- With 2:25 left in the same game and still 3 timeouts remaining, the Patriots faced 4th and 6 from the Broncos 14 trailing 20 to 12. Now what should they do?
- Belichick elected to go for it again, and they were again stopped short. This too was the call that maximized EWP, since $EWP(\text{field goal attempt}) = 0.147$ and $EWP(\text{go for it}) = 0.161$.
- Interestingly, the Patriots lost the game by 2 points, 20 to 18. If they had kicked a field goal in either one of the two fourth down situations, they might have won.

Real-time computing of EWP by coaches?

- NFL has computers on the sidelines and in the pressbox (<http://www.sfgate.com/technology/article/NFL-players-to-use-tablet-computers-during-games-5665371.php>), but it seems they are allowed to use these only to view static pictures of previous plays.
- College football doesn't yet allow computers for any in-game purpose (<http://www.timesunion.com/sports/college/article/Low-tech-College-football-yet-to-embrace-10914002.php>).