Michael Ortiz
David Kravitz
16 December 2011

Predicting NFL and NCAA Offensive Play Calling

Introduction

The NFL and NCAA Football are multi-billion dollar industry. Despite this, the process of preparing to face an opponent is a relative anachronism. Coaches watch and take notes of their opponents' game footage the week before game day. When it comes time to call plays, they rely on their memory and instincts in order to call the proper play. A defense would be at a distinct advantage if their coaches had reliable statistical data to complement their instincts. In this project, we use machine learning to predict the plays called by offensive coaches.

Building the data set

ESPN provides play-by-play data of NFL and NCAA football games. Here is an example:

http://espn.go.com/ncf/playbyplay?gameId=312460099

We wrote a program to fetch and parse this data into a computer readable format. For each play, we used the following features:

Input:
[home team, away team, home team score, away team score, clock time remaining, whether or not the home team is on offense, line of scrimmage, down, distance to go]

Output:
[play type (run or pass)]

NFL data set size: ~40,000 examples
NCAA data set size: ~30,000 examples

We ran separate classifiers for the NFL data and the NCAA data. We also ran classifiers on a per team basis, i.e. predicting play calls for just Dallas Cowboys, just the New England Patriots, etc.

Logistic regression for binary classification

Using k-fold cross validation (with k = 10), we ran a basic logistic regression classifier. At first, we used just the NFL data set and yielded a training set accuracy of 76% and a test set accuracy of 50%. This seemed to imply that we were suffering from a high bias problem. Thus, we tried adding a regularization term, which yielded a training set accuracy of 70% and a test set accuracy of 65%. To further improve our accuracy, we tried using polynomial features, i.e. adding features to the data set based on polynomial version of the original features. This did not help much, and we yielded similar training and test set accuracies as before. Here is a breakdown of the classification accuracy by data set:

Predicting run/pass for any NFL team: 65%
Predicting run/pass for a single NFL team: 61% (least-predictable teams) to 70% (most-predictable teams)
Predicting run/pass for any NCAA team: 70%
Predicting run/pass for a single NCAA team: 66% (least-predictable teams) to 73% (most-predictable teams)
We were interested to see how play calling predictability correlated to rush/pass ratio and win/loss percentage. The NFL is predominantly a passing league. For the 2005 NFL season, 44% of all plays were rushes and 56% were passes. Only three teams ran more rushing than passing plays in the season and the passing percentage varied from a low of 44% (Pittsburgh) to a high of 66% (Arizona).

We did not find it surprising that the most predictable team, Arizona, had the most skewed rush/pass ratio. They also had one of the worst season records at 5-11. New Orleans and Green Bay also had poor season records of 3-13 and 4-12, while over 60% of their plays called were passes. By contrast, the teams with the best records in 2005 were very balanced. Seattle, Indianapolis, and Denver all finished with 13 or more wins. Their passing percentages were, respectively, 50.6%, 54.7%, and 48.6%.

As mentioned earlier, NCAA teams proved to be more predictable than NFL teams by between 3% and 5%. The data suggests that this is because their rush/pass ratios skew away from 50/50. Many NCAA offenses tend to fall into a certain niche. Some teams, like Navy, call nearly 70% rushing plays, while teams like Houston, call nearly 70% passing plays. Play calling balance in the NCAA is a rarity and this fact likely contributes to the greater predictability of NCAA offensive play calling relative to the NFL.

SVM with linear kernel

Classification using an SVM with a linear kernel yielded similar results to using a logistic regression classifier.

SVM with RBF kernel and PCA for dimensionality reduction

SVM with an RBF kernel was too computationally expensive given the size of our data set. So, we used PCA to reduce the data to 2 dimensions. Surprisingly, our results were no better than logistic regression or a SVM with a linear kernel.

Recursive Feature Elimination

Given our SVM (with linear kernel) model, we used recursive feature elimination to find the most important features. Here are the features in order of importance:

Down
Distance to go
Whether or not the home team is on offense
# of plays that have occurred in the current period
Away team score
Period
Home team score

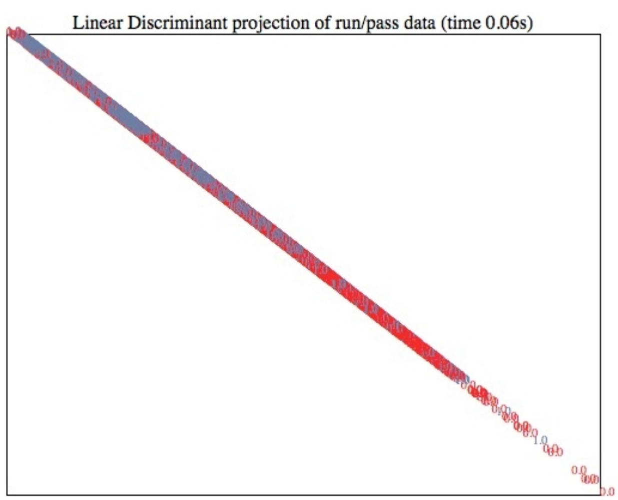
Home team
Line of scrimmage
Away team
Clock time remaining

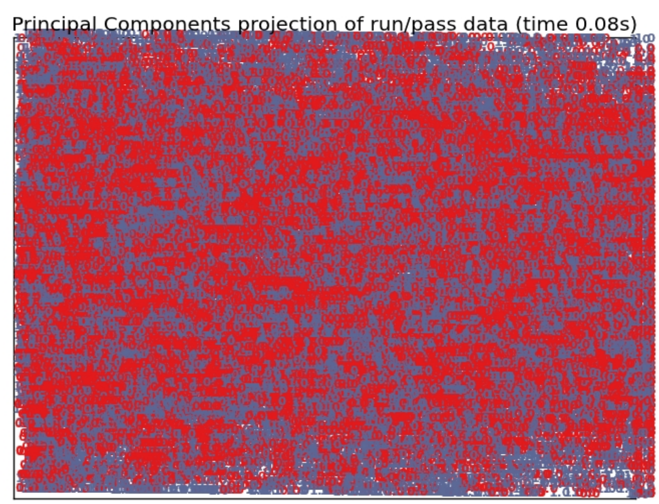
Predictably, down and distance to go are the most informative features. This is hardly news. A passing play typically averages more yards than a rushing play. So, when the distance to go to get the first down is still high, the play is more likely to be a pass. Conversely, when the distance to go is very short, it is more reliable to gain those yards with a rushing play.

Clock time remaining and line of scrimmage were our least informative features. This was also predictable. For the majority of the game time, the time remaining does not generally effect play calling. However, towards the end of the half or game, teams often pass more to gain yards quicker when they are losing. We included the clock time feature for this reason. Similarly, play calling is not affected greatly by line of scrimmage. However, when teams are backed up against their own goal line, they tend to rush. This is because a passing play can run the risk of a quarterback being tackled in his end zone. Such a play is called a safety, and rewards the opposing team with 2 points and possession.

2-dimensional PCA and LDA for data visualization

From the 2D plots it is clear that LDA outperforms PCA in terms of class discrimination. We know that LDA deals directly with the discrimination between classes while PCA does not take into account the underlying class structure. Thus, we know that the discriminatory information is not aligned with the direction of maximum variance.





Confidence in prediction

For the NFL data set, are there certain cases where we are very confident in our decision?

We can make a correct prediction with > 70% probability in ~18% of cases.

We can make a correct prediction with > 80% probability in ~6% of cases.
We can make a correct prediction with > 90% probability in ~1% of cases.

Where do the errors come from?

We had some success in predicting play calls, but we still suffered from large error rates. The best explanation for this is simply that coaches are not machines. There is an inherent randomness to their play-calling -- no coach follows an exact formula. Moreover, a good coach tries to surprise his opponent in order to gain an advantage.

Next steps and applications

One way we believe we could improve our predictions by including information about previous plays. For instance, if a team rushes twice in a row, are they less likely to rush a third time? If the previous pass was incomplete, are we more likely to see a rush or a pass?

Our ultimate ambition would be to be able to classify more detailed information about each play type than rush or pass. For instance, is the play a designed rush up the middle or a play action pass? This would involve a large time commitment to reviewing game film and classifying each play, which is a reason we decided not to pursue this implementation for this project. However, there is a chance that this data could be useful to coaches and is a logical next step.

The most obvious application of our research is as a tool to assist defensive play calling. If a defensive coach knows what type of play the opposing offense is likely to run, then they should be better prepared to defend against it. We were initially underwhelmed by our results. However, once put into perspective, 70% predictability of play calling should still be a useful tool for coaches.

In particular, coaches can cherry pick when they decide to follow our predictions based on the confidence of a given prediction. We predicted with over a 70% confidence level in approximately 18% of all plays. In a game where every yard matters, this kind of reliability for 18% of all plays can provide the edge to win many football games. Offensive coaches can also use it as a tool to discover their own predictable play calling biases, in order to become more unpredictable.