Cluster Analysis To Create Quarterback Archetypes

Gavin Byrnes

4 February 2016

Abstract

This paper introduces a multifaceted project to present a new non-linear approach to portfolio construction. Building a football team presents a suitable data set with convenient expert knowledge and a plausible simulation available. We then introduce a method of using cluster analysis and other devices to construct position archetypes, and apply it to the quarterback position as a test case for the rest of the team.

1 Introduction

One of the underlying goals of the Decision Lens Maths R&D Department is to improve on our current unrealistically linear approach to portfolio construction. Successful portfolios require a diverse variety of alternatives, and do not always have value conveniently equal to the sum of their parts. We have frequently used a football metaphor in making this argument, saying things like "Having ten good quarterbacks is not ten times as valuable as having one good quarterback" or "If you have a terrible quarterback, accumulating a lot of wide receivers doesn't help very much." We have constructed a far-reaching experiment on the mechanics and mathematics of portfolio construction based in the football metaphor.

1.1 Why Football?

- Building a football team is a cost-constrained activity that requires analyzing and evaluating alternatives (players) both individually and as a team on a series of criteria both qualitative (scouting reports, intangibles) and quantitative. (physical measurements, on-field performance)

- The success of a football team is not determined by the mere sum of the overall talent of its players; there are certain types of players that complement each other better than other types would and certain roster constructions that allow the team to get by with expending fewer resources elsewhere, e.g. a team with a great offensive line can probably get good production from a relatively inexpensive running back and use the savings elsewhere rather than spending more money on an expensive running back.[1]

- It is of course impossible in a real world portfolio evaluation to choose several ideas and examine alternate universes. Football is not perfectly predictable, but the simulation abilities of the Madden video game series allow us to evaluate different team constructions in a reasonably accurate environment.

- Madden also has very detailed player ratings, assigning a score between 1 and 99 to every player on 52 different attributes, from tackling to short pass accuracy to kick power. Ratings collection for criteria is as simple as downloading a publicly available spreadsheet.

We will therefore be analyzing football roster construction, using (at first) the ratings and simulation engine of the Madden video game series as our baseline set of data.

Before proceeding with the experiment, we need to introduce the idea of player archetypes and the cluster analysis we will be using to discover them.

1.2 The Importance of Cluster Analysis And Player Archetypes

In the introduction, we touched on the idea that some positions create important synergies with each other that are valuable to understand in constructing

[1] For a recent example, consider the Dallas Cowboys. Their excellent offensive line helped running back DeMarco Murray to a star quality season in 2014, after which he was signed to an expensive contract by the rival Philadelphia Eagles. The Eagles' offensive line performed poorly in 2015, leading to a less effective running game; in contrast, Dallas got solid performance out of much cheaper running backs because of their stellar line. Of course, it didn't matter in the end when Tony Romo got injured...

rosters. We can go a step further and note that there are valuable differences among players at the same position.[2] For example, some quarterbacks are among the worst athletes on their teams, but are also the most valuable players on their teams because of their incredible poise and accuracy (Tom Brady, Peyton Manning until recently). Some quarterbacks rely on speed and rushing ability to make up for some shortcomings in the more traditional passing areas (Michael Vick, Johnny Manziel). Some quarterbacks are less one-dimensional, whether they are very strong runners who can also pass well (Cam Newton), very strong passers who are still a minor threat to run (Aaron Rodgers, or somewhere in between. (Russell Wilson)

In this paper and future papers, we will be using cluster analysis to go a step beyond merely saying how "good" a player is to also evaluate how well he fits a particular archetype at his position. This will in turn allow us to examine nuances of roster construction that could allow for more optimal resource allocation. For example, suppose that a team is choosing between two quarterbacks. Quarterback A is better overall, mostly because he has a much stronger arm, though Quarterback B is a little more mobile. Not surprisingly, he is also more expensive. However, the team has a relatively poor offensive line and their wide receivers catch balls well but do not have gamebreaking speed. Quarterback A's great arm looks fine on paper, but if his slower receivers don't have time to get down the field before the porous offensive line lets in defenders, the team might actually be better off (or at least not that much worse off) with Quarterback B, whose greater elusiveness may help him avoid sacks and get the ball out to closer receivers! They can then use the money saved by opting for B over A to upgrade the offensive line or the defense. We know that portfolio construction is more complicated than simply adding up value and running VROI analyses; we hope that this experiment will help us quantify that understanding better in this instance and provide a springboard to a generalizable nonlinear heuristic for portfolio construction that will add value to Decision Lens in the future.

1.3 Aims Of This Paper

While the eventual goal of our project is to reach useful conclusions about roster-building, we need to lay important groundwork first; demonstrating

[2] And, by the same token, differences among any portfolio's alternatives that fit one particular category

that we can perform adequate cluster and archetype analysis at one position, generalizing this process to the other positions, and finally bringing the positions together for roster analysis. This paper and its companion present the work we did to get from raw Madden ratings on quarterbacks to clusters and archetypes, with a new grade assigned to each player based on how well he fits each archetype.

2 Data Collection

The data for this project are the ratings for the video game Madden 2016, found at [**?**].

1. We imported the ratings for quarterbacks into R.
2. We subjectively reduced the number of criteria from 52 to 21, based on the observation that some abilities (Tackling, Catching, Kick Power, etc.) are irrelevant to playing quarterback
3. We also included the Overall rating calculated in Madden to use as a sanity check and a guide to the relative importance of different criteria.[3]

3 Processes

The first thing we did was to cluster the quarterbacks based on their ratings on these 21 categories. This was useful as an extremely basic sanity check on the clustering algorithm; generally speaking, good quarterbacks clustered with other good quarterbacks, scrambling quarterbacks clustered with other scrambling quarterbacks, etc. However, this was insufficient for several reasons:

1. We want the archetypes of quarterbacks to be things like "pocket passer" and "running QB", not "good quarterback" and "bad quarterback"; this would ruin the point of comparing and contrasting across archetypes.

[3] Using expert knowledge in this way is something we wanted to avoid, but decided that we could not. For more details, see the companion paper (write it, then cite it).

2. Clustering on raw attribute scores tends to put good quarterbacks with good quarterbacks and bad quarterbacks with bad quarterbacks, even if their *relative* strengths and weaknesses don't match up. Therefore, we want to normalize scores when we cluster players. (but not when we rate them, of course)

3. The easiest and most intuitive way to normalize the scores is to make every player's ratings add up to the same number. However, the ranges of values differ greatly in both mean and standard deviation. For example, most quarterbacks have an Injury rating between 85 and 95 and a Throw Accuracy Deep rating between 50 and 75, while their Trucking ratings range from 20 to 76!

4. The mathematically intuitive way to ensure that the normalization of ability is not dominated by the more variable attributes is to first convert the scores to z-scores.[4]

5. Finally, having so many attributes makes the clusters awkward to visualize. In addition, anecdotal evidence suggests that some attributes, like Speed and Acceleration, or Throw Accuracy Short and Throw Accuracy Medium, would be very highly correlated. We decided that it would be a good idea to create clusters of the attributes themselves as well as the players to construct a smaller number of metafeatures to use in the eventual player clustering.

Therefore, we performed the following process:

- Begin with the data frame of the overall ratings and 21 raw attribute scores for quarterbacks, QBOriginal.

- Construct the data frame QBzscores, obtained by replacing each element with its zscore using the means and standard deviations of each column of player ratings.

- Construct the data frame QBNormzscores, which removes the Overall score and then adjusts every player's z-scores so that every player's sum of 21 attribute normalized z-scores is 0.

[4] Calculated for each data point by subtracting the mean of the overall dataset and dividing by the standard deviation

- Run a multiple regression to express the Overall z-score as a function of the un-normalized attribute z-scores. This will help us provide a baseline weight for the attributes when we construct archetypes, as, even among running quarterbacks, throwing ability is still the most important consideration.[5]

At this point, we diverged down two potential paths, both of which follow the same steps with a slightly different data set. In one case, we performed a covariance and cluster analysis on our 21 attributes to group them in smaller metafeatures, followed by proceeding with the rest of our analysis. In another case, we first used the multiple regression we ran against the Overall score to remove categories below a certain threshold of relevance, and then performed the same analysis on the 10 attributes that remained.

3.1 Without Criteria Reduction

Without a criteria reduction, we entered this phase of the process with the following data frames:

- QBzscores: Z-scores on the Overall score and 21 attributes
- QBnormzscores: Z-scores on 21 attributes, adjusted so that each player row sums to 0.
- covar2: The covariance matrix for all 21 attributes, based on relatively correlated player performance on pairs of attributes (i.e., players with high scores on Speed tend to have high scores on Acceleration)

We performed a cluster analysis on the covariance matrix covar2, emerging with five clusters:

- Running QB: Acceleration, Speed, Carrying, Throw on the Run
- Awareness: Play Action, Awareness
- Throwing: Throw Accuracy Short, Throw Accuracy Mid, Throw Accuracy Deep, Toughness, Strength, Throw Power, Injury, Stamina

[5]Facts like this contribute to why we were unable to avoid using the overall score in our process. More on this in the other paper

- Fast Run: Spin Move, Juke Movie, Ball Carrier Vision, Elusiveness
- Power Run: Agility, Trucking, Stiff Arm

We used these meta-features to create a new dataset, QBCluster2, with 6 columns (Overall, Running QB, Awareness, Throwing, Fast Run, Power Run). The individual attributes were weighted within their own meta-features based on their coefficients in the multiple regression with Overall. If the coefficient was negative they were assigned zero weight. We then re-normalized these results to QBNormCluster2 by removing the Overall element and once again setting all row sums to zero.

QBNormClusters2 is the dataset we had been working towards: players were normalized so that ability should not be a signficant factor, they had a manageable number of scores, and their relative strengths and weaknesses in each of the five meta-features had been captured. We were therefore able to perform a cluster analysis on the players, called QBNormkcluster2. We experimented with a few numbers of clusters before settling on 5, resulting in the following means and descriptions: table

1. Scrambler: High normalized scores on Power Run, Fast Run, Running QB. Low scores on Throwing and especially Awareness.
2. Low Awareness Balanced: Slightly low Awareness, around average or slightly above average on everything else.
3. West Coast: High Awareness, High Throwing, Slightly low Power Run, slightly low Running QB, Very low Fast Run.
4. Pocket Passer: Very High Awareness, High Throwing, Low Fast Run, Very Low Power Run, Very Low Running QB
5. High Awareness Balanced: Slightly High Awareness, Slightly Low Power Run, average everything else.

With these archetypes constructed, we now had to score every player's un-normalized attributes against these archetypes. We did this as follows:

1. Run a multiple regression to express the Overall score in terms of the five meta-features from QBCluster2, normalizing the resulting weights to sum to 1. These weights are known as standardunreducedweights2,

and they, unsurprisingly, suggest that Awareness and Throwing are far more important to the average quarterback than their various expressions of running ability.

2. Convert the cluster means for each archetype into raw cluster weights of each meta-feature, producing the data frame clusterunreducedweights2. To do this, we took the pnorm (cumulative density function) of each value and normalized so that each row summed to 1. For example, the cluster weights for Scrambler were constructed by taking the pnorms of (0.43, -1.09, -0.71, 0.67, 0.69) to get raw weights of (0.67, 0.14, 0.24, 0.75, 0.76) and then normalized to a sum of one for the cluster weights of (0.26, 0.05, 0.09, 0.29, 0.30).[6]

3. Now armed with both standardunreducedweights2 (an understanding of the average contribution of the meta-features to quarterback quality) and clusterunreducedweights2 (an understanding of the meta-features relatively prioritized by each archetype), we construct adjustedunreducedweights2 by setting each archetype's weight to be the average of its cluster weight and the standard weight.[7]

4. Finally, we apply the standard unreduced weights to the un-normalized meta-feature scores QBCluster2. The result is a number in z-score format, which we then took the pnorm of and multiplied by 100 to place on a 0 to 100 scale to produce the final score set QBTotalScoresUnreduced2. The average quarterback on a given meta-feature should score a 50, with the caveat that with 119 quarterbacks in the data set and only 32 starting spots available in the NFL, in general quarterbacks are skewed towards a very large group of mediocre to bad ones with a few standouts.

3.2 With Criteria Reduction

With a criteria reduction, we used the multiple regression against overall score to eliminate some attributes that did not significantly contribute to

[6]These numbers are rounded, of course; the actual calculations include many more digits

[7]If we didn't do this, we would end up with undesirable situations in which fast but inaccurate quarterbacks are highly overrated. More on this in the other paper.

the overall score. This left us with ten known useful attributes: Speed, Acceleration, Agility, Awareness, Throw Power, Throw Accuracy Short, Throw Accuracy Mid, Throw Accuracy Deep, Play Action, and Throw On The Run. We reduced the data set to just the overall z-score and these 10 attributes, re-normalized to have the row sums equal 0, and entered this phase of the process with the following data frames:

- QBReduced: Z-scores on the Overall score and 10 attributes
- QBNormReduced: Z-scores on 10 attributes, adjusted so that each player row sums to 0.
- covar: The covariance matrix for the 10 attributes, based on relatively correlated player performance on pairs of attributes

We performed a cluster analysis on the covariance matrix covar, emerging with 3 clusters as follows:

- Scrambling: Speed, Acceleration, Agility
- Accuracy: Awareness, Throw Accuracy Short, Throw Accuracy Mid, Throw Accuracy Deep, Play Action
- DeepBall: Throw Power, Throw on the Run

We used these meta-features to create a new dataset, QBCluster, with 4 columns (Overall, Scrambling, Accuracy, DeepBall). The individual attributes were weighted within their own meta-features based on their coefficients in the multiple regression with Overall. If the coefficient was negative they were assigned zero weight. We then re-normalized these results to QBNormCluster by removing the Overall element and once again setting all row sums to zero.

We were therefore able to perform a cluster analysis on the players, called QBNormkcluster. As in the other analysis, we ended with 5 clusters, resulting in the following means and descriptions: table

1. Scrambler: Very High Scrambling, Very Low Accuracy, Low DeepBall
2. Mobile Deep Thrower: Average Scrambling, Low Accuracy, High DeepBall

3. West Coast: Average Scrambling, High Accuracy, Low DeepBall

4. Big Arm: Very Low Scrambling, Average Accuracy, Very High DeepBall

5. Pocket Passer: Very Low Scrambling, Very High Accuracy, High DeepBall

With these archetypes constructed, we now had to score every player's un-normalized attributes against these archetypes. We did this as follows:

1. Run a multiple regression to express the Overall score in terms of the five meta-features from QBCluster, normalizing the resulting weights to sum to 1. These weights are known as standardreducedweights.

2. Convert the cluster means for each archetype into raw cluster weights of each meta-feature, producing the data frame clusterreducedweights. To do this, we took the pnorm (cumulative density function) of each value and normalized so that each row summed to 1.

3. Now armed with both standardreducedweights and clusterreducedweights, we construct adjustedreducedweights by setting each archetype's weight to be the average of its cluster weight and the standard weight.

4. Finally, we apply the standard unreduced weights to the un-normalized meta-feature scores QBCluster. The result is a number in z-score format, which we then took the pnorm of and multiplied by 100 to place on a 0 to 100 scale. The resulting data set is the overall score set, QBTotalScoresReduced. The average quarterback on a given meta-feature should score a 50.

4 Results and Tables

Maybe have this just refer to the RMarkdown document, can I insert it in here?

5 Analysis and Sanity Check

A sanity check of our overall results should include two things: first, good quarterbacks should still be (generally) good and bad quarterbacks should be (generally) bad. Second, and more importantly, we should be able to see particularly specialized quarterbacks doing better relative to the archetypes they fit and worse relative to those they don't. We analyze this by running correlations between archetype scores and the original Overall score and by examining players with particularly high or low scores on certain archetypes relative to others.

5.1 Unreduced Analysis

The five archetypes of Scrambler, Low Awareness Balanced, West Coast, Pocket Passer, and High Awareness Balanced have the following correlation coefficients with the Overall score:

Make a table of this: 0.69 0.84 0.97 0.98 0.91

This makes sense; The quarterbacks closest to the West Coast and Pocket Passer archetypes are the most traditionally successful, and their strengths are those categories that most contribute to a high overall score. This doesn't mean that it is bad to be a good scrambling quarterback; it just means that it is easier to be a good scrambling quarterback but a bad quarterback overall (or a bad scrambling quarterback but a good quarterback overall) than it is to be a good pocket quarterback but a bad quarterback overall (or a bad pocket quarterback but a good quarterback overall)

Comparing individual scores, the first thing to note is that our new archetype scores range from the teens to the high 90s, while the Overall rating scores bottom out in the mid 50s. This, of course, is due to us only considering quarterbacks; Madden allows you to play non-quarterbacks at quarterback (it is not a good idea), so no real quarterback can be below the 50s as the potential 10s and 20s of quarterbacking play need to be set aside for the linemen and running backs of the world.

The second thing to note is that a lot of this sanity check will be anecdotal, examining individual alternatives and thinking "this placement makes sense" or "this placement does not make sense." This is useful to know for future applications; while we intend to improve on portfolio construction, we will not produce a solution that removes the need for expert evaluation and participation. In this case, I am a sufficient expert on the NFL in general and Madden in particular to perform a reasoned sanity check.[8]

5.1.1 Great Runners, Terrible Runners

An easily recognizable quarterback archetype here are the fast runners who aren't very good at passing. Players like Robert Griffin III, Johnny Manziel, and Tyrod Taylor (before the 2015 season) are high up in the Scrambler rankings but not very strong overall. On the opposite end of the spectrum are talented but very slow quarterbacks like Carson Palmer, Philip Rivers, and Tom Brady. On the Pocket Passer archetype, Tom Brady is a 94 and Joe Webb a 40...on the Scrambler archetype, they are both 66s.

5.1.2 Do Players' Strongest Archetype Match Their Cluster?

Another thing to check is, does each player's highest archetype score occur on the archetype to which he was assigned in clustering? If not, is it at least close? I was slightly dissatisfied with the results in this case; there were no egregious errors, but it was very common for a player who was clustered into the Low Awareness Balanced or High Awareness Balanced to have his highest archetype score come as either a Scrambler or a Pocket Passer. Probably this is because those are extreme archetypes; Pocket Passer did have the highest standard deviation of scores of the five archetypes, and many quarterbacks in the Low Awareness Balanced cluster are mediocre passers who would benefit from the Scrambler archetype's relatively lower emphasis on passing ability.

I would consider the Unreduced analysis a qualified success; no egregious errors, and some useful information, but possibly too much irrelevant information entering into the clustering.

5.2 Reduced Analysis

The five archetypes of Scrambler, Mobile Deep Thrower, West Coast, Big Arm, and Pocket Passer have the following correlation coefficients with the Overall score:

Make a table of this: 0.87 0.93 0.97 0.95 0.98

The correlations with overall rating are similar to those in the unreduced analysis, with Pocket Passer and West Coast highest and Scrambler lowest.

[8]For reference, see my life from age 13 to 17

We notice that the correlation is stronger when we took out attributes that contributed less to Overall rating.

5.2.1 Dominance Of The Pure Quarterback

Removing attributes with lower correlations with overall rating leads unsurprisingly to stronger results for more traditionally excellent quarterbacks. While Cam Newton ranks as the best Scrambler archetype in the Unreduced analysis, his Reduced score is behind Aaron Rodgers and Andrew Luck. Whereas in the Unreduced analysis mediocre overall quarterbacks jumped up the Scrambler list quite quickly, here even the slowest star quarterbacks don't fall too far down the list; for example, Tom Brady's 75 on Scrambler archetype still beats the likes of EJ Manuel and Johnny Manziel.

In general, there is less variability when the attributes are reduced and trimmed to only the most important to quarterbacking. As the correlations with Overall rating imply, there is precious little in the way of deviation and even known particular strengths and weaknesses might not change much. To give an example, going into the 2015 season Peyton Manning's overall Madden rating was 92 and Matthew Stafford's was 84. Matthew Stafford has one of the strongest arms in football (in fact, he is the only player in the game with a 99 Throw Power), while Peyton Manning's once strong arm has been severely weakened by repeated neck surgeries. Despite this, Peyton scores an 86 on the Big Arm archetype to Stafford's 85.

Like the unreduced analysis, the reduced analysis doesn't say anything outlandish, which is important; it also has more intuitive clustering (Big Arm, etc.) than the rather vague-r Awareness-based clustering of the unreduced analysis. On the other hand, the reduced analysis may suffer from being too much of a recreation of the overall rating process. Our eventual portfolio construction plan relies on testing the idea that a "worse" player may be a better team fit than a "better" one because of his greater skills in a particular area; if the archetypal differences are weak enough that not much differentiation occurs, this may be harder to measure and explore.

5.3 Comparison Of Methods

The reduced and unreduced methods we have used here both have strengths and weaknesses relative to one another, though they both have succeeded in being reasonable overall. The reduced method has more intuitive clustering and better correlation with the overall rating; the unreduced method allows for real variance to stand out more effectively. The unreduced method is probably preferable for the continuation of the overall project, but it may be worth some investigation into tweaking it a little further.

5.4 Clustering As Player Prediction

Though it is outside the scope of this project, we should note that this type of clustering analysis could be useful in evaluation of young players or amateur prospects. When skill level is normalized out of the equation, clustering could provide young players role models to shoot for; players who have the same relative strengths and weaknesses that they do, who are older and better and represent an aspiration.

6 Future Research

We have established a method for beginning with a set of raw attribute ratings and an overall expert assessment and ending with a set of archetypal ratings derived from cluster analysis that takes into account covariance between individual attributes and similar makeups of alternatives. The next step is to apply the analysis across other positions on a football team and, from there, to experiment with different methods of team construction using both real examples (where applicable) and Madden simulations.

We will also be on the lookout for more efficient simulations or for more effective ways to reduce how many different kinds of teams we need to construct; a full analysis of even a smallish fraction of the amount of teams we could plausibly create might take too long to be valuable.

List Of Variables

References