**Grit Analysis Research**

**My Motivation**

My original idea for the project came from an interest in the concept of grit and whether it could be quantified. Knowing that I would not have access to players personality evaluations, I thought about how I reasonably predict a player’s personality traits with public data. I stumbled on a paper by Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner from the University of Maryland, entitled “Predicting Personality from Twitter.” The paper detailed how certain language features found in tweets are significantly correlated with the Big 5 personality traits: extraversion, conscientiousness, neuroticism, openness, and agreeableness. In these Big 5 personality traits, I thought there may be some connection with what we in the baseball community would call grit. Grit in my view is not about the raw talent, but more of a testament to the players’ ability to improve oneself.

A close up of text on a white background

Description automatically generated

The chart showing correlation between language features and certain personality traits. The bolded numbers indicate a significant correlation.

**Hypothesis**

Because conscientiousness is often associated with job success, I thought there may be a correlation between conscientiousness and grit (or improvement).

**Selection of Sample**

My sample was made of players who made their rookie debut in 2014 or 2015 according to Fangraphs. There were over 30 players in the original sample, but when I filtered for seasons where players had 400 PA, certain players like Travis Jankowski no longer qualified. I did not use any pitchers in my sample.

**How I Quantified Improvement**

I wanted a well-rounded statistic and despite the problems with WAR, I thought it was probably the closest to a holistic indicator of skill that is currently available. I used Baseball Reference data, so WAR is calculated using their formula. I thought that the difference between the first season and the most recent season would not be a full picture of improvement, especially if the first or most recent season were outliers compared to other performance. Instead, I averaged the difference between consecutive seasons. (ex. 2016 WAR – 2015 WAR, 2015 WAR – 2014 WAR, etc. averaged)

**Selection of Language Features**

For the project, I selected only some of the language features shown in the paper because I wanted to simplify the search process and ensure the greatest accuracy. My work includes 8 features in total, a combination of the punctuation features, “You”, and Non-LIWC features. See the full list with their associated personality traits below.

A screenshot of a cell phone

Description automatically generated

Full list of metrics with their associated personality traits

**Process**

I began by combining Lahman’s database with baseball reference data. Once I had cleaned up the data, using the process outlined by Kris Eberwein in R-bloggers, I created a CSV file with the combined data. I needed to get my dataset ready next with my chosen sample of players, their twitter handles, their average WAR growth (my chosen improvement metric), and an average of their ranking for each personality trait. I started by completed a vlookup of sorts to link the Fangraphs player id with the baseball reference player ID through the Lahman master file. The function I used allowed me to join my list of players found in Fangraphs with the baseball reference data.

Once I had my chosen players and WAR data in a database, I calculated the average growth as the mean of the difference between two consecutive seasons. (for more information, see How I Quantified Improvement) I then added a list of Twitter accounts to create the first half of my dataset (Player, Average WAR growth, and Twitter account handles)

In order to organize my process, I included a matrix which shows the list of all the language features I had chosen to use with the associated personality traits and whether the metrics were positively or negatively correlated. (See the list above)

Next, I created a Twitter developer account in order to conduct research with Twitter data. The account was free, but if this research was to be repeated you would need to go through the process as well. With my account information, I was able to use the rtweet package (linked in Sources) to database 500 of the most recent tweets from each player in my sample.

After completing my database of tweets, I had to find how often each language feature was used in each player’s tweets, filtering out retweets. I divided the total count of language features used by the total number of tweets, so each language figure count is on a per tweet basis. For example, I found all of the parentheses use in tweets then divided the number by the total tweets (excluding retweets) by the total tweets of each player. I then created a ranking of which players had the highest to lowest total usage per tweet. I created a ranking system because I wanted to be able to average the rankings to get a general sense of which players were high in certain personality traits and which were low. There were also multiple language features for some personality traits used, so I needed to ensure I could combine my results to create one score for each player’s personality traits.

After I averaged the rankings for each language feature associated with each personality trait, I had a score for personality trait for each player. Finally, using a pearson test of correlation, I found the correlation between the average WAR growth and each personality trait score.

**Results**

I found no significant correlation (p = < .10) using the ranking method and comparing the data to improvement. The correlation was between -.30 and .37 with Openness being the lowest value and agreeableness being the highest. The probability values were all greater than .10, but the openness score probability value and agreeableness score probability value were the closest to being significant. The sample size in the end was 27. My results do not indicate that my hypothesis is correct but suggest there may be more work needed on the subject.

**A screenshot of a cell phone

Description automatically generated**

**Future Iterations**

I completed multiple iterations of the project including one where I did not rank the players based on use of language features and instead looked at what features may have correlated with improvement. I did not find any significant results but will put my work in my Github folder. I also tried to see if any language features correlated with a player’s average WAR to see if there was a correlation between personality traits and ability. I found a significant correlation between WAR and word count, which could probably be explored further.

In the future, some changes could be made to the research to increase the likelihood of a significant outcome. I would increase the sample size in order to increase the likelihood of significance. I would also look into other possible metrics besides WAR; other metrics may show better results. I am not convinced that my method of calculating improvement was the optimal approach. I would be open to looking into other options. Further, I would take a different approach with my ranking process because it may not have been the best way to estimate a player’s personality. I could also increase the number of tweets used to increase the dataset. Another possible iteration of the project could filter out players who did not have enough tweets, but I didn’t want to limit my sample size even further.

Sources

<https://www.demenzemedicinagenerale.net/pdf/2011%20-%20Predicting%20Personality%20from%20Twitter.pdf>

<https://cran.r-project.org/web/packages/rtweet/rtweet.pdf>

<https://www.r-bloggers.com/how-Krito-add-war-metrics-to-your-lahman-database/>

<https://www.baseball-reference.com/>

<https://www.fangraphs.com/>