Spencer Tollefson

November 5, 2018

Proposal for Project Fletcher

LeBron James' Popularity on reddit.com/r/nba

Introduction

reddit.com/r/nba is the subreddit dedicated to the National Basketball Association (NBA). With over 1,420,000 subscribers at the time of writing - its membership dwarfs all other major American sports leagues' subreddits in size and even the 1,217,000 subscriber-strong r/soccer. Subreddit submissions cover everything from potential recruits playing in the NCAA and for international teams, game threads for all NBA games played, breaking NBA news, speculative questions by redditors, light-hearted jokes, and everything in between.

Unsurprisingly, one of the most well-known players currently playing in the NBA, LeBron James, is the direct or an indirect focus of many forum submissions. His popularity and the general NBA fandom's sentiment of him has knowingly ebbed and flowed over time. For example, when he announced he was leaving the Cleveland Cavaliers on a live television special in 2010 to play in Miami, the general sense was people viewed him negatively as it was perceived he was "teaming up" with other good players. On the other hand, he has had moments where the public has viewed him in an overwhelmingly positive light, such as when he contributed to building an elementary school in summer 2018.

The **goal** of this project will be to perform a sentiment analysis of r/nba's viewpoint on LeBron James over a period of years. It will be fascinating to see if the subreddit 's mood corresponds to real life events involving LeBron. Other NLP techniques will be employed to try and decipher more detail surrounding LeBron.

Question/Need

How has LeBron James' popularity changed over time? Is it tied to specific subjects/reasons?

This question will be explored by performing NLP techniques to obtain general sentiment of r/nba users feelings toward LeBron James.

Additional potential questions:

- Are there trends when LeBron is more or less popular? (I.e. during the playoffs, after his team loses, if he has a viral tweet, etc)
- Does the language of affirmers or detractors change over time? For example, did negative posts in 2008 accuse him of "selfishness" while in 2017 it accused him of "sharing the ball" too much?

Data

I plan to use the reddit API to obtain all comments above a certain vote-threshold from posts with "LeBron James" or a variation of his name within the title. The dataset will likely be quite large, as my intuition estimates have been multiple posts about LeBron daily in the subreddit for years. Using Sentiment Analysis I will score the polarity (pos/neu/neg) of each post.

I plan to eventually create a table like the following:

comment_text	time	reddit_score	reddit_user	pos_score	neu_score	neg_score	compound_score
string	datetime	int	string	float	float	float	float

Additionally, I am considering pulling other sources of data such as game results and tweets to bring in additional information. Twitter could be used to compare sentiment vs reddit r/nba sentiment.

Characteristics of each row of data

Each row of table will contain 1 comment and its attributes.

Known Unknowns

- What type of big question I will be trying to answer. My biggest interest right now is seeing if sentiment analysis will map popularity to real life events
- It is unclear how well sentiment analysis will detect inside jokes and intentional / unintentional grammar mistakes unique to r/nba