PIGSKIN

Visualizing Football Tweets

TEAM: Andrew Chao (@acchao) • Gilbert Hernandez (@thegilby) • Jacob Portnoff (@jacobportnoff)
MENTOR: Gilad Mishne (@gilad)

Goals

- Successfully distinguish between tweets about professional football & those that are not
- Visualize tweets about football in a meaningful way
- Analyze & compare tweets in relation to particular teams, match-ups, & our dataset as a whole

Approach

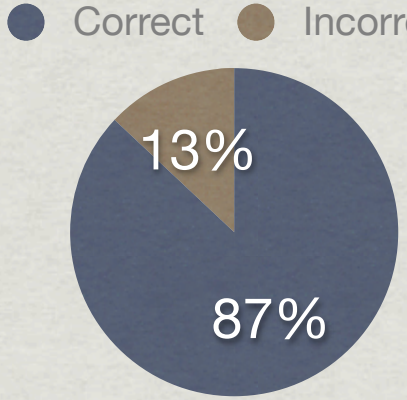
- Research Preferred Methods
- Data Acquisition
- Data Analysis
- Data Visualization

Data Acquisition

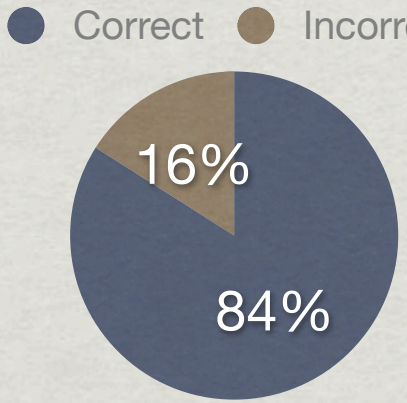
- Collecting tweets using unfiltered Twitter stream on Sundays from 10AM - 9PM Pacific
- ~2 Million tweets / day
  - Data collection error left us w/ only partial data for Nov 18 (~1.5M tweets)
- ~14 Million tweets in our dataset at the moment
- Tweet data collected: date, time, username, tweet, hashtags, & geolocation
- Indexed tweets using Apache Solr

Algorithm Development

FOOTBALL TWEET RECALL



FOOTBALL TWEET PRECISION



CURRENT WEIGHTING RESULTS

Lists used:

Term list

- Generated by online football dictionary
- Very common terms removed or improved
  - (Down -> 1st down, first down, etc.)

Team list

- 32 teams
- Some variations (cowboy, cowboys, cowboyz)

@usernames

- 32 generated from verified accounts

Player names

- ESPN player list (Full name only)

Hashtags

- Top Hashtags from verified team accounts

Nicknames

- Nicknames from Wikipedia dataset
- Weighted depending on commonality

Difference from previous lists (w/ nicknames, improved term)

Recall (how clean): steady

- 71% -> 87%

Precision (how complete): nearly doubled precision

- 57% -> 84%

Data Visualization

- Using d3.js & Rickshaw visualization libraries
- Total volume of (non) football tweets by week
- Tweet rate during each matchup
- Top tags found each week / matchup


DEMO
http://groups.ischool.berkeley.edu/pigskin/


END
Thank you! Comments & Questions?
Andrew Chao (@acchao) • Gilbert Hernandez (@thegilby) • Jacob Portnoff (@jacobportnoff)
http://groups.ischool.berkeley.edu/pigskin/