Finding Stock Exposure from Twitter

Nick Tyler

Overview

- Use tweets from popular business accounts to find topics that move stocks
- Map what is being tweeted about to a particular stock and how the price moves
- Use this mapping to gain insights into the underlying factors that go into a stocks price

Project Outline

1. Fetch the data for tweets and stocks
2. Clean and process the data
3. Create models to predict if a tweet is relevant to a particular stock
4. Model business tweets and measure the movements of stocks relative to tweet topics

Dataset

- Daily stock data from AlphaVantage (closing prices)
- 30 Dow stocks are the domain
- Twitter API
  - Tweets from popular business users back to the beginning of time
  - “Random” tweets are samples from the past 7 days

Data Processing

- Day over day prices used to determine stock movements
- Text processing of tweets into records:
  - Stopping
  - Stemming
  - Tokenization

Tweet Mapping Model

- Tweets pulled from hashtags of company names and acronyms
- Random tweets pulled from twitter to represent the negative class
- Build predictive models to determine if a tweet is related to a stock
- Individual models per stock
- Logistic, Bayesian, and Ensemble models used.

Tweet Mapping Model (cont)

Naive Bayes: 85.4%

Logistic: 85.3%

ADABoost: 85.4%

Random Forest: 85.4%

Only run analysis on stocks where a model with accuracy > 80% was achieved

Results of Final Model

- Use linear model to determine the topic vectors that were most important to the movement of a stock
- Map the topics back to individual words to make them human decipherable

Results of Final Model (cont)

KO (Coke):

plastic: 14.4967

doctor: 14.01

pollut: 12.0854

world: 7.9427

driver: 7.01

PFE (Pfizer):

inhibitor: 0.01

result: 0.01

roll: 0.01

stage: 0.01

clinic: 0.01

WBA (Walgreens):

boycott: 116.01

patient: 7.4469

deni: 1.0146

medic: 1.01

commit: 1.01

Thank You

Questions?