Using Social Media Data as Early Warning Signals in Risk Management

Joanne, Helen and Luna

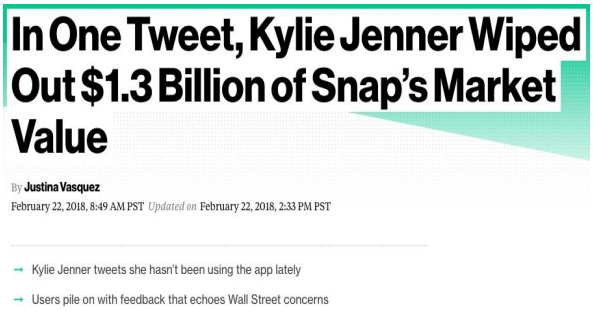
Recap: Question & Motivation

Can social media data sentiment analysis be used to predict stock trends? Do mentions of companies affect its stock value?

Puppy dies after United flight attendant forces it to ride in overhead bin

Share / Tweet / Reddit / Flipboard / Email

EDITOR'S NOTE: The dog that died on a United Airlines flight barked for two hours in the overhead bin. Read the latest story here.



In One Tweet, Kylie Jenner Wiped Out $1.3 Billion of Snap's Market Value

By Justina Vasquez

February 22, 2018, 8:49 AM PST *Updated on* February 22, 2018, 2:33 PM PST

- Kylie Jenner tweets she hasn't been using the app lately
- Users pile on with feedback that echoes Wall Street concerns

Data

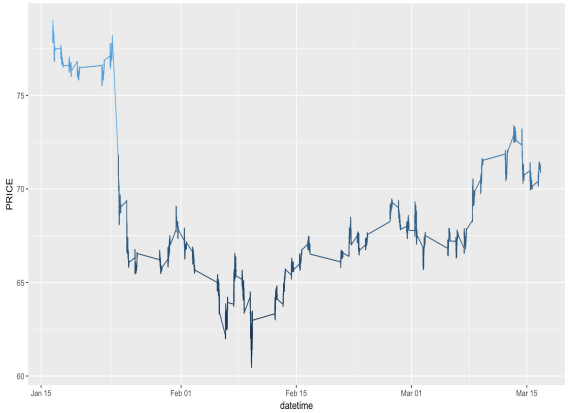
- Twitter tweets directed @united
  - Twitter API
  - 175,000+ instances
  - March 7, 2018 -- March 16, 2018
- Intraday (high frequency) UAL Trade Prices
  - Wharton Database Research Services, The Wharton School at the University of Pennsylvania
  - UC Berkeley Haas School of Business
  - 1,444,514 instances
  - January 16 , 2018 -- March 16, 2018 (9:30 am -- 4:00 pm)

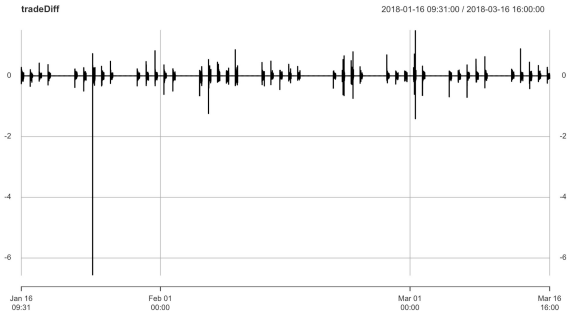
Literature References

- Twitter sentiment analysis
  - Using Twitter as a source of information for time series prediction (Ramon Xuriguera, Universitat Politècnica de Catalunya)
  - The sentimentr Package (https://cran.r-project.org/web/packages/sentimentr/README.html)
- Time series of high frequency intraday
  - The highfrequency Package (to aggregate secondly trade into every minute)
- Intervention analysis
  - Time Series Intervention Analysis with R and SAS (Matt Bogard, econometricsense.blogspot)
  - STAT 510 9.2 Intervention Analysis (PennState)

1. Use the data before the intervention point to determine the ARIMA model for the series.
2. Use that ARIMA model to forecast values for the period after the intervention.
3. Calculate the differences between actual values after the intervention and the forecasted values.
4. Examine the differences in step 3 to determine a model for the intervention effect.

EDA and Plots






```
adf.test(tradeAgg, alternative = "stationary") # check stationary p-value = 0.5136
d = ndiffs(tradeAgg) # difference = 1
tradeDiff <- diff(tradeAgg, differences = d)
tradeDiff[1] = tradeDiff[2]
adf.test(tradeDiff, alternative = "stationary")
```

Detailed Question - Intervention Analysis

- Find intervention point from sentiment analysis
- Split stock data by intervention point
- Fit model on training set (data before intervention point)
- Forecast data values after intervention point
- Figure out intervention effect by comparing with true values after intervention point
- Determine a model for the intervention effect from last step
- Get overall model including intervention effect

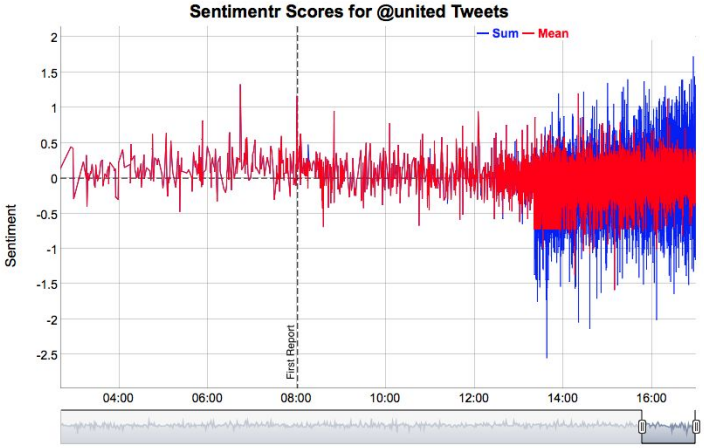
Preliminary Result - Sentiment Analysis

Results from using sentimentr package

Interactive plot using dygraphs package

Tweets are recorded in UTC time

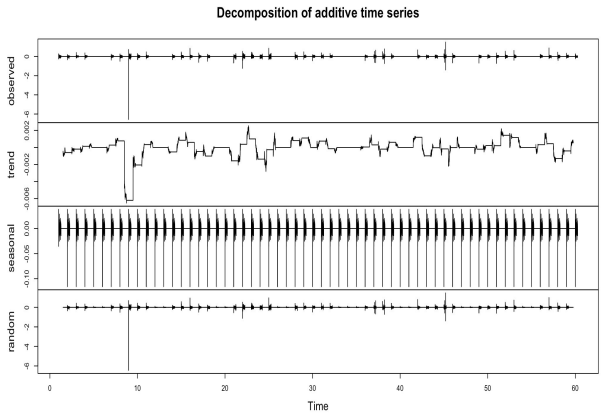
Intervention point:
2018-03-13 15:02:00 PST



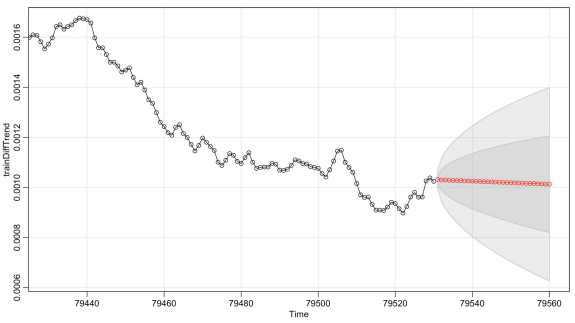
Opinion lexicon English textfiles
sentimentr uses a much faster regex based approach that is nearly as accurate in parsing sentences with a much lower computational time. We see that **RSentiment** and Stanford take the longest time while **sentimentr** and **syuzhet** are comparable depending upon lexicon used.
It is a dictionary lookup approach that tries to incorporate weighting for valence shifters (negation and amplifiers/deamplifiers).

Rudimentary: opinion-lexicon-english a list of positive and negative words

Trend Component



sarima.for(trainDiffTrend, 30, 1, 0, 4)

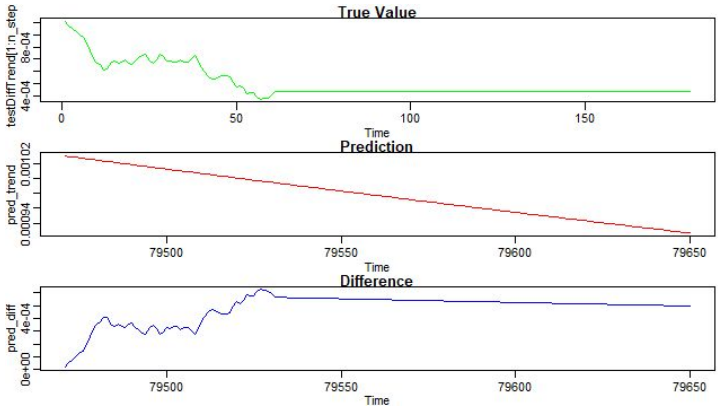


Adjust seasonal
tradeDiffSeasonal <- tradeDiff2 - tradeDiffDecomp$seasonal
Use trend to de-seasonal

TSA: Start to Finish Examples (Ryan Kelly June 19, 2014)
Before using sarima model, we compared with Forecasts using Exponential Smoothing (HoltWinters), ets (models error, trend and seasonal elements together). Arima lowest AIC

sarima.for(trainDiffTrend, 30, 1, 0, 4)

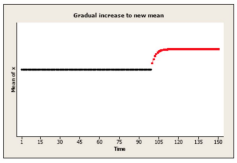
Preliminary Results - arima



Intervention point:
2018-03-13 15:02:00

Intervention Pattern:

Gradual increase or decrease to a new mean level:
There may be a gradually increasing amount that is added (or subtracted) which eventually levels off at a new level (compared to the "before" level).



Pattern of intervention

1. Permanent constant change to the mean level: An amount has been added (or subtracted) to each value after time T.
2. Brief constant change to the mean level: There may be a temporary change for one or more periods, after which there is no effect of the intervention.
3. Gradual increase or decrease to a new mean level: There may be a gradually increasing amount that is added (or subtracted) which eventually levels off at a new level (compared to the "before" level).
4. Initial change followed by gradual return to the no change: There may be an immediate change to the values of the series, but the amount added or subtracted to each value after time T approaches 0 over time.

Used arima(1,0,4) to forecast values after intervention point "2018-03-13 15:00:00". Plot the difference between true value and prediction. Notice that the pattern looks like gradually increasing which eventually get a new mean level.

Future Work

- Continue pulling twitter data for lasting sentiment analysis
- Better model to fit stock dataset
- Intervention analysis (pattern, model)
- Regression using sentiment scores and tweet information as features

Let z_t = the amount of change at time t that is attributable to the intervention. By definition, z_t = 0 before time T (time of the intervention). The value of z_t may or may not be 0 after time T.

Add a vector or matrix of external regressors like Libor did in his slides