E-Ship: Modeling public and entrepreneur sentiment from longitudinal online data

Johan Bollen¹, Krishna Bathina¹, and Timothy Slaper²

(1) School of Informatics, Engineering, and Computing Center for Complex Networks and Systems Research (Complex Systems)

(2) Indiana Business Research Center, Indiana University

Objectives

E-Ship:

- Compound indicator of regional development and resilience
- "traditional" regional indicators
- "unconventional" indicators





Economic data

- Growth
- Employment
- Trade

Surveys:

- Consumer sentiment
- Entrepreneur ship
- Funding

Unconventional data

- Social media
- Search volumes
- Sentiment
- Network data

Approach

Measuring population-wide sentiment over time and validating it against major events, regional indicators, and entrepreneurial activity and resilience.

E-Ship relevant work:

- 1. Measuring collective population-level sentiment from online data at the county-level for US
- 2. Cross-validation against major events and disasters (possibility of measuring resilience, time to recovery)
- 3. Cross-validation against large-surveys of Subjective Well-Being at state-level in the US

Diagram of Data and Processing



Our data

- 3,221 US counties US Census Bureau 2010
- Mean 3434.8 Tweets/county
- Median 327 Tweets/count



Sentiment Analysis:

Mapping text content to numerical indication of sentiment

Natural Language Processing:

- 1. Algorithm parses text
- 2. Knowledge base, lexicons, classifier -> sentiment classification or rating

"Today SUX!"

- Negative: 0.779
- Neutral: 0.221
- Positive: 0.0
- Compound: -0.5461

"Johan is smart, handsome, and funny."

- Negative: 0.0
- Neutral: 0.254
- Positive: 0.746
- Compound: 0.8316

Challenges:

- Accuracy: does rating reflect actual sentiment?
- Internet lingo, abbreviations, short blurbs, emojis, creative use of punctuation
- We use VADER (Hutto & Gilbert, 2014):
- Open-source
- Rated highest performing in large-scale survey

Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014. Social media specific capabilities

1. Punctuation:

That was great vs. That was great**!!!!**

2. Capitalization: That was great vs That was GREAT

3. Degree Modifiers:

That was great vs. That was **really** great

4. Contrastive Conjunction:

That was fun **but** I didn't like it **1. Trigram** analysis to find negation That was **not that great**

Face validity and drawbacks of VADER

Extreme negative outliers:

- "David and Nick are the 2 most disgusting leaders this country has ever seen, neglecting working families and making us carry the nation"
- "Gedaffi sons daughters and grandchildren will censured en and grandchildren burn them alive destroy the evil pastarus"
- "Gedaffi your daughter and sons will die die slow painfull deaths, your slag daughter will bε

But there are some false positives:

Bot-generated horoscope content for *Cancer*.

- "You may be pleased at how creative you can be if you simply gi... More for Cancer<u>http://t.co/FFo9KTCP</u>"
- "You would be a happy camper if you could stay at home today, b... More for Cancer<u>http://t.co/FFo9KTCP</u>"
- "You may be attracted to an emotional drama today because the e... More for Cancer<u>http://t.co/FFo9KTCP</u>"

• .

This may require further advances in sentiment analysis, but VADER was shown to outperform most tools with very high accuracy in recent survey, see Ribeiro et al (2015)

Twitter data clean up and time series construction

- Clean-up of Tweet content:
 - Remove state acronyms with VADER sentiment values (NH, TX, OK)
 - 2. Remove hyperlinks
- Calculate sentiment user VADER (or custom Vader), each tweet -> [-1,+1]
- Group tweets by week

For each week:

- 1. 95% confidence interval from bootstrapping average Twitter valence estimation
- 2. Null-model comparison: random tweets with same weekday distribution as given week

US County-level sentiment mapping tool



jbollen@i







Complete county-level sentiment data is now available in easy-to-parse JSON format

as well as

County-specific sentiment time series with null-model and 95% CIs

Case studies for validation: face validity

 We chose 3 locations hit by hurricanes

validation: do we find a
Twitter sentiment signal at
that time?
Resilience:

Does sentiment return to baselines?

- How fast?

Houston, TX: "Harvey", Aug 17 -Sep 22, 2017 2. Puerto Rico (entire territory): "Maria", Sep 20, 2017 3. Florida (entire state): "Irma", sep 10, 2017







E-Ship validation efforts (summer 2018)

- 1. Investigation of ability to detect sentiment specifically for online communities of entrepreneurs on Twitter
- 2. Cross-validation against state-level survey data of Subjective Well-Being in USA

Comparing sentiment: Entrepreneurs v. general population

Are we truly measuring entrepreneurial sentiment via Twitter?

Data:

Entrepreneur data

- 251 entrepreneur user accounts
- Each with up to 3,200 most recent tweets
- Tweet count distribution:

Null-model: Random tweets from one random day

- We chose 2015-03-23
- Over 36 million tweets

Null-Model: Random user timelines

- N=251
- Over 15,000 tweets
- Tweet count distribution:



Technique 1: VADER

Method

- Rate each "entrepreneur tweet" using VADER compound score
- Compare entrepreneur tweets to a day's worth of random tweets
- Because we do see positive skew at the tweet level for entrepreneurs, a potential userlevel score would be the VADER mean of all authored tweets



Technique 2: Opinion Finder

Instead of evaluating all text sentiment, we rate "subjective", i.e. personal statements only

<u>Method</u>

- OpinionFinder (OF) was developed to focus specifically on subjective sentiment
- For each entrepreneur timeline, we gather positively- and negatively-classified words by OF
- Entrepreneur individual sentiment score:
 - score = (#positive #negative)/(#positive + #negative)
 - Collect only if number of classified words in the timeline is > 10



OF random timeline scores

Cross-validation Twitter vs. state-level SWB surveys:

Are we measuring Subjective Well-Being or another population sentiment via Twitter?

Behavioral Risk Factor Surveillance System

- 1. CDC nationwide telephone survey about health and risk behavior
- 2. Started in 1984 with 15 states
- 3. Over 400,000 participants per year
- 4. Subjective well being measure of quality of life Confirmation (Owald et al.)
 - a. Presence of positive emotion
 - b. Absence of negative emotions
 - c. Life satisfaction
 - d. Fulfillment
 - e. Positive functioning

Objective Confirmation of Subjective Measures of Human Well-Being: Evidence from the U.S.A.

SWB question

Label: Satisfaction with life

Section Name: Emotional Support and Life Satisfaction

Section Number: 24

Question Number: 2

Column: 699

Type of Variable: Num

SAS Variable Name: LSATISFY

Question Prologue:

Question: In general, how satisfied are you with your life?

Value	Value Label	Frequency	Percentage	Weighted Percentage
1	Very satisfied	14,600	49.45	47.04
2	Satisfied	13,240	44.85	46.50
3	Dissatisfied	1,127	3.82	4.42
4	Very dissatisfied	327	1.11	1.31
7	Don't know/Not sure	130	0.44	0.47
9	Refused	99	0.34	0.26
BLANK	Not asked or Missing	456,780	•	•

Survey Results

- 1. Survey Range 2005 to 2016
- 2. Twitter started in 2006 data available afterwards
- 3. Very incomplete data not all questions mandated by all states
- 4. Later years especially only have data on few states

Subjective Well Being vs Sentiment

- Do they correlate for each state? - hard to check because the later years are not complete
- 2. Used the mean of states over all years

no significant correlation at this point



Behavioral Risk Factor Surveillance System 2

- 1. Gallup Sharecare Well Being Index 2017
- 2. 2.5 million surveys
- 3. 5 dimensions
 - 1. Purpose Liking what you do every day
 - 2. Social Relationships
 - 3. Financial Economic life
 - 4. Community Liking where you live
 - 5. Physical Having good Health

Subjective Well Being vs Sentiment - State

r squared = -0.101

no significant correlation at this point



Subjective Well Being vs Sentiment - County

r squared = -0.05

no significant correlation at this point



Continuing work 2018-2019

- Refinement of sentiment analysis tool with objective to increase validity with respect to SWB and other social indicators
- Applications of different sentiment analysis indicators: "text sentiment" vs. "personal SWB"
- Measuring regional and entrepreneurial resilience from longitudinal data
- Contributing validated sentiment data to E-Ship construction in collaboration with other partners