

Sentiment Analysis

Atlanta-area Evaluation Association
Monday, February 22, 2014

Emily Baughman



Sentiment Analysis



Turning attitudes and opinions in useful data



The material presented reflects my own personal view and not that of any organization.

Sentiment? Analysis?



Learning Objectives

- ☞ Understand sentiment analysis/ opinion mining
- ☞ Know the tools and resources available
- ☞ Identify ways to use sentiment analysis techniques for your next evaluation



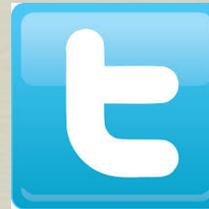
Overview



- ✧ **Online Behavior:** What are the trends around social networking?
- ✧ **Sentiment Analysis/Opinion Mining:** What is it?
- ✧ **How-to:** What are the tools and skills needed?
- ✧ **Evaluation Applications:** How can sentiment analysis fit into an evaluation strategy?

Internet Usage

- 86% of U. S. adults and 95% of teens use the internet
- 73% of U.S. adults who are online use a social networking site; 42% use multiple social media sites
- Common social media sites: Facebook (71%), LinkedIn (22%), Pinterest (21%), Twitter (18%), Instagram (17%)



Online Behavior

- ∞ Rise in “shareable content” and “viral content”
 - ∞ Facebook: **360 million** Facebook users update their Facebook status at least once a week, **125 million** users update their status at least once a day.¹
- ∞ Specialized communities
 - ∞ E.g., health, parenting, private business
- ∞ Online commentary and reviews



Bottega Louie
★★★★☆ **7426 reviews** Details
\$\$ · Italian, Bakeries, Breakfast & Brunch Edit



treatments
The Case Against Multivitamins Grows Stronger
by NANCY SHUTE
December 17, 2013 12:37 PM
Share
Comments

Blogs

- Over **409 million people** view more than **14.4 billion pages** each month.¹
- Users produce about **40.5 million new posts** and **50.3 million new comments** each month.¹
- Microblogging: blogging with space constraints
 - Twitter, Facebook and LinkedIn status updates

AEA365 | A Tip-a-Day by and for Evaluators

[Home](#) [About](#) [Contribution Guidelines](#) [Archive](#)



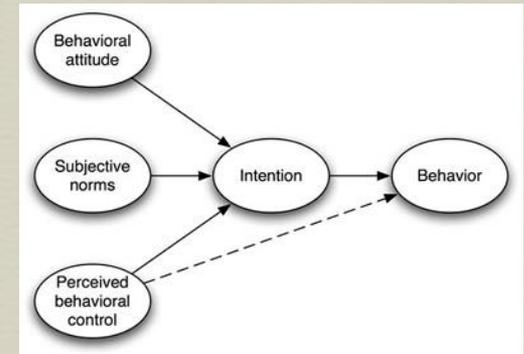
Twitter



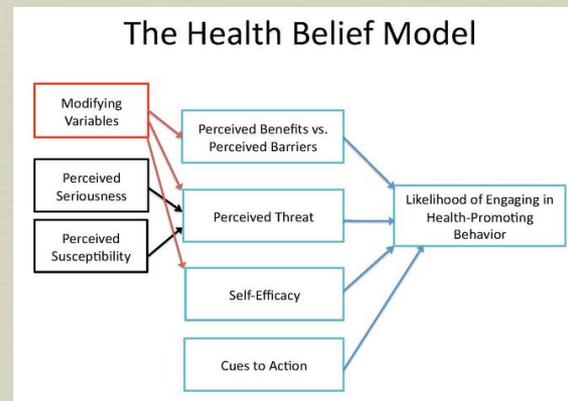
-
- ∞ 18% of internet users¹
 - ∞ Adults ages 18-29, African-Americans, urban residents
 - ∞ 241 million monthly active users²
 - ∞ 250 million tweets per day³

Why do we care about attitudes?

∞ Theory of Planned Behavior



∞ Health Belief Model



∞ Transtheoretical Model



Solving Offline Problems through Online Networks

∞ Crowdsourcing



∞ Purchasing behavior influenced by online intentions



∞ 57% of Super Bowl 2014 commercials included a hashtag (#)



Sentiment Analysis / Opinion Mining

- ☞ Computer Science meets Social Science
- ☞ Classifying opinions into quantifiable data
- ☞ 😊 / ☹ or happy, angry, sad
- ☞ Common applications: product and brand marketing, politics, finance (behavioral economics)

Sentiment Analysis in the News

Enterprise

big data

thomson reuters

Twitter

Thomson Reuters Taps Into Twitter For Big Data Sentiment Analysis

Posted Feb 3, 2014 by [Ingrid Lunden \(@ingridlunden\)](#)

6

Like

325

Tweet

1,152

Share

355

'Analytics 3.0 is about speed and scale': Prof. Thomas Davenport

by [Brian Pereira](#), [InformationWeek](#), February 13, 2014

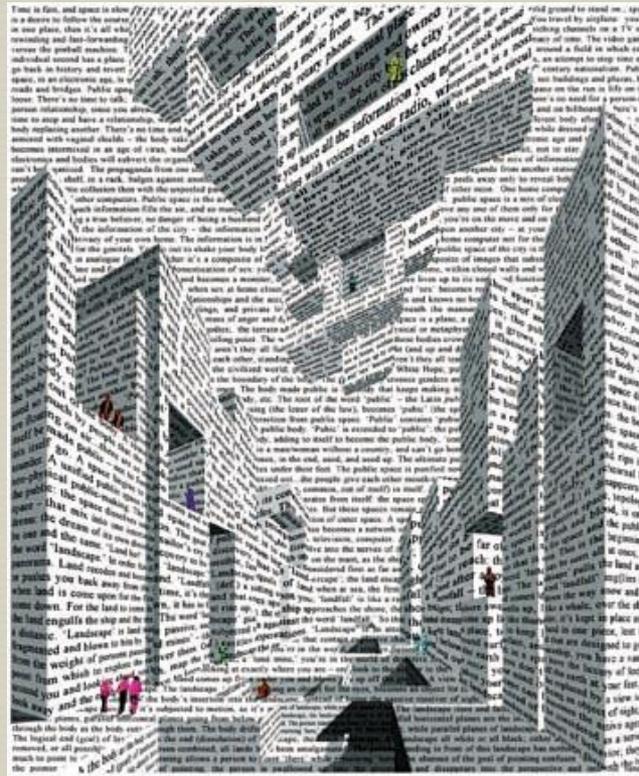
Sentiment Analysis As Customer Service

CUSTOMER EXPERIENCE

By [Aaron Mandelbaum](#), Published January 30, 2014

Be the first to comment!

Big Data



“City of Words” lithograph by Vito Acconci, 1999

Data, Information, Knowledge Pyramid

∞ Data:

∞ “raining”

∞ Information:

∞ “It is raining now”

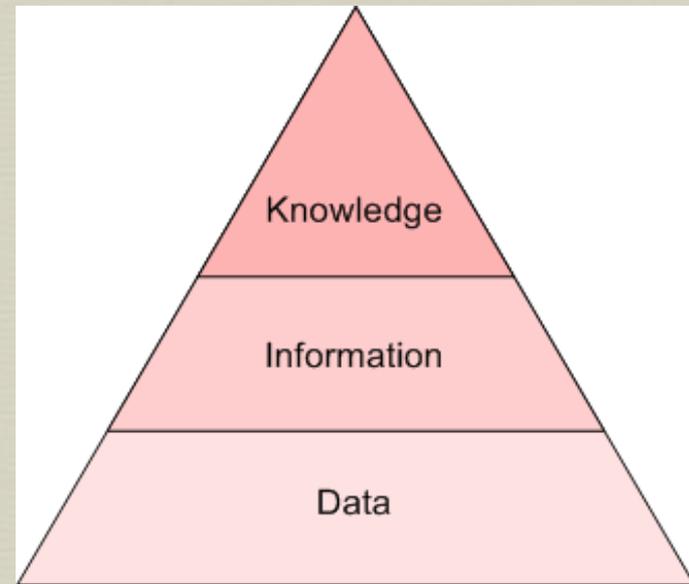
∞ “It won’t stop raining”

∞ “It is raining outside”

∞ Knowledge:

∞ “It is often raining in Seattle”

∞ “It is raining so I should bring my umbrella”



Sentiment Analysis “by hand”

Goal: Classify five tweets as positive, negative, or neutral.

Question: How do Twitter users feel about breastfeeding?



askthelc: To swaddle or not to swaddle, that is the question. #newborns #breastfeeding

Posted: 22 minutes ago

To swaddle or not to swaddle, that is the question
#newborns #breastfeeding



Doctorsbfc: **#Breastfeeding** decreases baby's risk of ear infections.

Posted: 35 minutes ago

#Breastfeeding decreases baby's risk of ear infections



[alyssar4890](#): @haileynhellums omggg those ppl are still arguing over the **breastfeeding!!!** They all have wayyy too much time on their hands lol

Posted: 1 hour ago

Omggg those ppl are still arguing over the breastfeeding!!! They all have wayyy too much time on their hands lol



CBerbs: @ilandgrl It's frustrating to have the same conversation again and again, but important for both mothers and babies. **#breastfeeding**

Posted: 2 hours ago

It's frustrating to have the same conversation again and again, but important for both mothers and babies **#breastfeeding**



IBIMA: New Study: Breastfeeding Counselling can be a Matter of Interprofessional Communication
<http://t.co/8L4JDMnQ49>

Posted: 2 hours ago

New Study: Breastfeeding Counselling can be a
Matter of Interprofessional Communication

How did you make your decision?

- ∞ # of positive or negative words?
- ∞ general impression of the sentence?
- ∞ implied meaning?
- ∞ some combination of techniques?

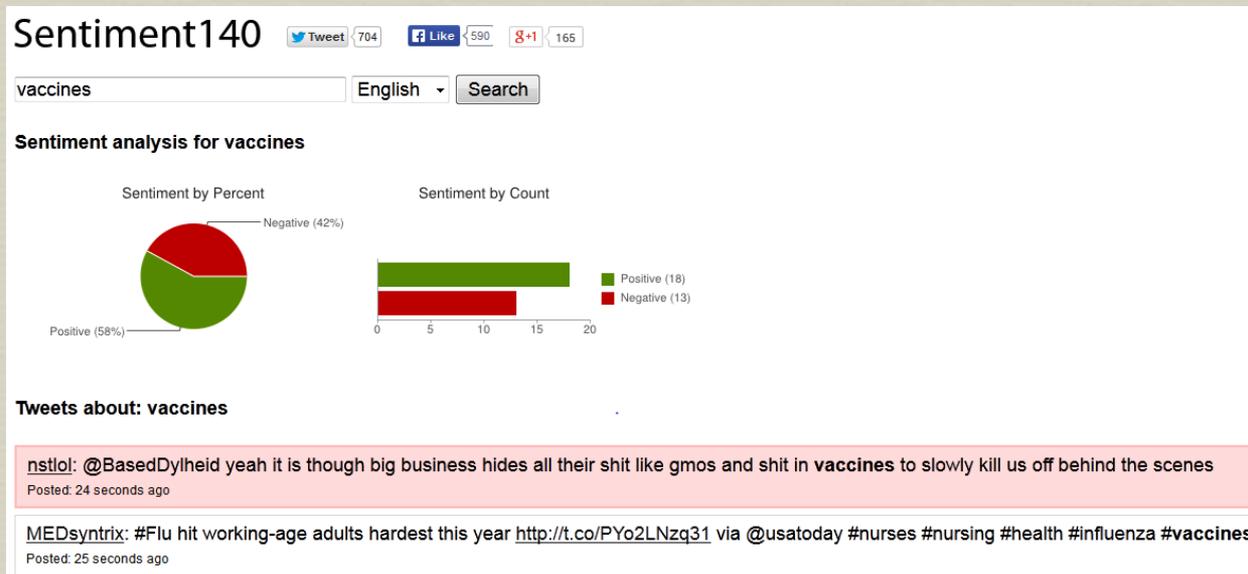
Sentiment Analysis Methodologies



- ∞ Document, Sentence, Feature/Aspect level
- ∞ Frequency of terms
- ∞ Part-of-speech (e.g., adjective, noun)
- ∞ Semantics (e.g., wonderful / terrible)
- ∞ Sentiment shifter (e.g., not / not only...but also)

Case Study: Sentiment140

🔗 www.sentiment140.com; “Vaccines”



Sentiment140 Methodology:

<http://cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>

Case Study: Sentiment140

TinaTinde: RT @PascalBarollier: @ONECampaign this proverb applies to the @GAVIAlliance 2 & **vaccines** are part of the solution. <http://t.co/VzjeQV7eUo> h?
Posted: 31 minutes ago

Vivekkrishna11: RT @arrahman: Thanks to millions of vaccinators & millions of **#vaccines** #India is polio-free Now let's **#endpolio** everywhere <http://t.co/z1L> ?
Posted: 32 minutes ago

CUMedicalSchool: RT @ChildrensColo: **#Vaccines** are a hot topic with strong opinions on both sides. See how our experts answered your questions: <http://t.co/D> ?
Posted: 32 minutes ago

davidtlander: My blog on classic thrillers that influenced Vaccine Nation **#thrillers #action #suspense #autism #vaccines** <http://t.co/eWUliPGzGs>
Posted: 37 minutes ago

nstlol: @BasedDylheid yeah it is though big business hides all their shit like gmos and shit in **vaccines** to slowly kill us off behind the scenes
Posted: 45 minutes ago

EyesOnFreight: RT @UPS_Canada: Ever wonder how **vaccines** and medical supplies get from point A to point B? <http://t.co/uU6Y89ZtaJ>
Posted: 1 hour ago

chel_eeeeee: But i am sooo tired of seeing posts about **vaccines** and being unvaccinated. Rant ovaaa.
Posted: 23 minutes ago

soaked2thebone: Are we still doing **#unpopularopinion** thing? I normally support **vaccines** but am skeptical of the flu shot. It's a crap shoot.
Posted: 27 minutes ago

TCNavigator: Don't worry all - no polio, yellow fever or typhoid for me now. **#vaccines #international #travel**
Posted: 28 minutes ago



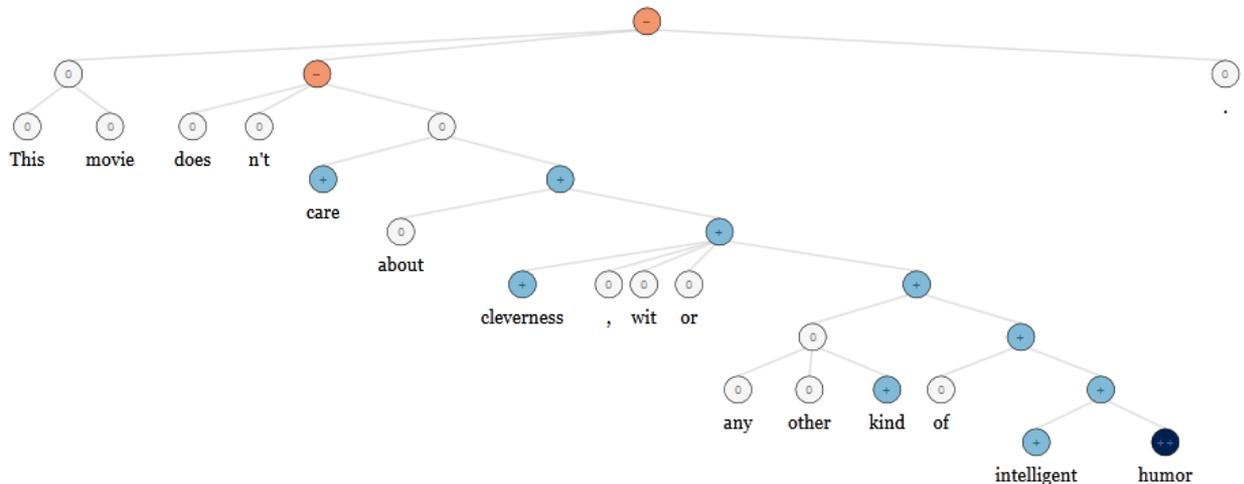
Case Study: NLP Movie Reviews

Stanford sentiment analysis modeling demo: <http://www-nlp.stanford.edu/sentiment/>

“This movie doesn’t care about cleverness, wit or any other kind of intelligent humor.”

Sentiment Trees

You can double-click on each tree figure to see its expanded version with greater details. There are 5 classes of sentiment classification: **very negative**, **negative**, neutral, **positive**, and **very positive**.



All labels are now correct

Natural Language Processing

∞ Analyzing and interpreting human speech

∞ Semantics, syntax

∞ Decision tress, ontologies

∞ Susan Bennett as Apple's 'Siri'



Machine Learning Algorithms

- ∞ Neural network based on weighted data
- ∞ Sorting, patterning data based on human feedback
- ∞ Examples:
 - ∞ Spam filters
 - ∞ “Personalized” music playlists
 - ∞ Product recommendations

Use Amazon Betterizer to Like Items

Let us know about the items you like to help personalize your shopping experience on Amazon by getting new product recommendations instantly.

Validation with Empirical Data

∞ Public Health

- ∞ Twitter sentiments on flu vaccine strongly correlated ($r = .78$) with flu vaccination rates by geographic region (Salathe & Khandelwal, 2011).
- ∞ Google flu searches strongly correlated ($r = .72$) with CDC Virus Surveillance (Ortiz et al., 2011).

∞ Politics

- ∞ Twitter sentiment predictive of German election results based on number of mentions (Tumasjan et al., 2011).

∞ Finance

- ∞ Twitter moods used to predict daily up and down changes in the Dow Jones Industrial Average with 87.6% accuracy (Bollen, Mao, & Zeng, 2011).

Sources: Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS computational biology*, 7(10), e1002199.

Ortiz, J. R., Zhou, H., Shay, D. K., Neuzil, K. M., Fowlkes, A. L., & Goss, C. H. (2011). Monitoring influenza activity in the United States: a comparison of traditional surveillance systems with Google Flu Trends. *PloS one*, 6(4), e18687.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with Twitter how 140 characters reflect the political landscape. *Social Science Computer Review*, 29(4), 402-418.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.

Limitations

- ❧ No algorithm is perfect
 - ❧ Solution: Opinions are subjective by nature
- ❧ Undetectable linguistic nuances (e.g., sarcasm)
 - ❧ Solution: Classify by hand a sample
- ❧ Limited ability to segment audiences
 - ❧ Solution: Target social platforms and online communities frequented by your audience

Software

∞ Paid

∞ Radian 6

∞ Crimson Hexagon

∞ Trackur

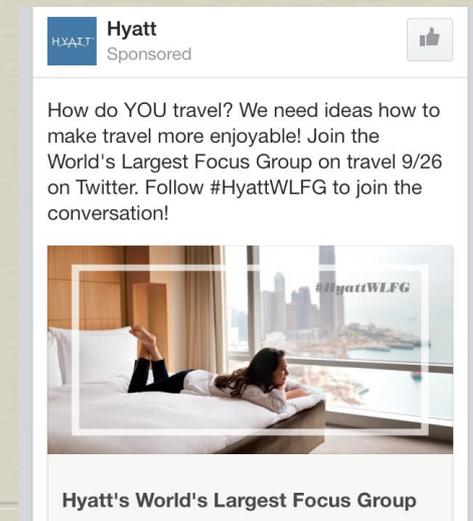
∞ Free

∞ Twendz

∞ SocialMention

Evaluation

- ❧ Low cost
- ❧ Creative use of metrics
 - ❧ Keyword search: Are people searching for information about the evaluated program?
- ❧ More observational?
- ❧ Help to explain evaluation results
- ❧ Twitter Focus Groups



The image shows a screenshot of a tweet from Hyatt, marked as 'Sponsored'. The tweet text reads: 'How do YOU travel? We need ideas how to make travel more enjoyable! Join the World's Largest Focus Group on travel 9/26 on Twitter. Follow #HyattWLFG to join the conversation!'. Below the text is a photograph of a woman lying on a bed in a hotel room, looking out a large window at a city skyline. The hashtag #HyattWLFG is overlaid on the photo. At the bottom of the tweet, it says 'Hyatt's World's Largest Focus Group'.

Rethinking Data Gathering

∞ Old way: Data Collection

- ∞ Interviewers, surveys, participant incentives, response forms

∞ New way: Data Mining

- ∞ Machine learning, data discovery, algorithms, models, automated



Understand your metrics

- ❧ Website **hits** = “**how idiots track success**”
- ❧ Engagement, Reach, Influencers
- ❧ Validate metrics with survey responses



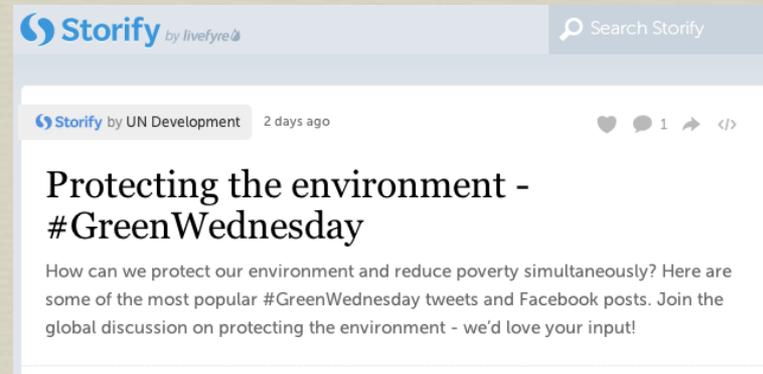
Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted. - Albert Einstein

Make Social Media Tell a Story

Storify (www.storify.com)



A sustainable agriculture can not only create opportunity for future environmental development but also eradicate poverty in the rural area.



The Future?

↪ Moore's Law: "computing power" doubles every two years

↪ Emoticons? 😊 😞



↪ Intention/ Behavior Analysis?



Additional Resources

❧ Books and articles

❧ Sentiment Analysis and Opinion Mining (2012) Bing Liu

❧ Articles by Guru, Seth Grimes (<http://altaplana.com/grimes.html>)

❧ Coursera Courses

❧ Machine Learning (March 3):
<https://www.coursera.org/course/ml>

❧ Practical Machine Learning (Jun
2):<https://www.coursera.org/course/predmachlearn>

Question and Answer



“The important thing is not to stop questioning.” – Albert Einstein

Thank you!

Emily Baughman
emilyabaughman@gmail.com