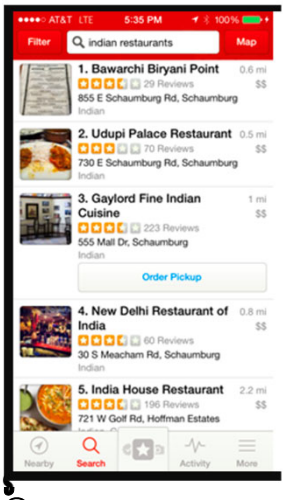
<Your Name>

Sentiment Mining

Sumeet Kumar
sumeetku@cs.cmu.edu

The CASOS Center
School of Computer Science, Carnegie Mellon
Summer Institute 2019

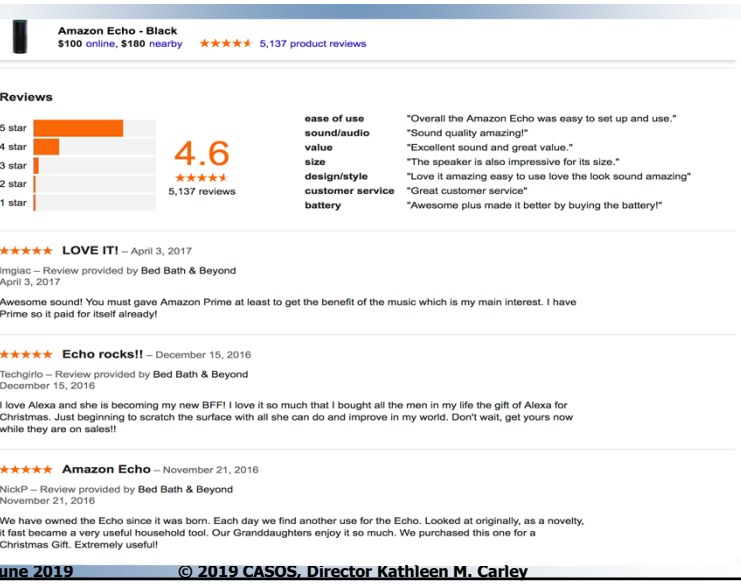
Why Sentiment Analysis?



<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Why Sentiment Analysis?

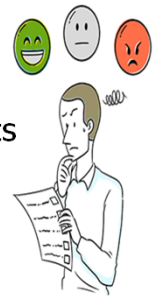


Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

What is Sentiment Analysis?

Goals:

- Elicit emotional responses in internet exchange
- Attitude of the writer towards a topic
- Often outputs polarity (-ve, +ve, neutral) or scale (1,..., 5)



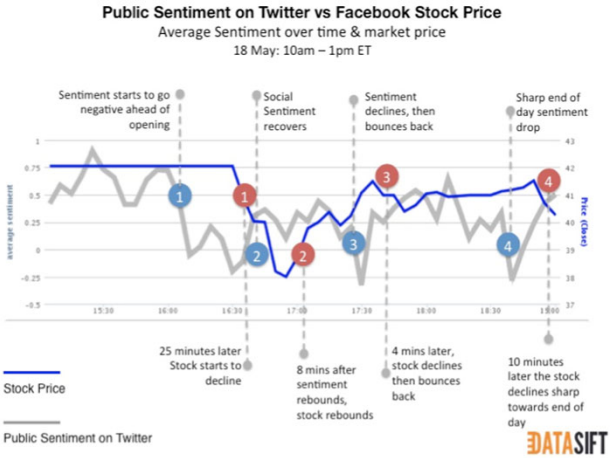
Use Cases:

- Business looking to market their products
- Understanding voters
- Build Networks??

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Use Case: Linking Text Sentiment to Stock Market



Source: http://socialmediaimpact.com/5-reasons-twitter-becoming-essential-stock-market/

Carnegie Mellon
institute for SOFTWARE RESEARCH

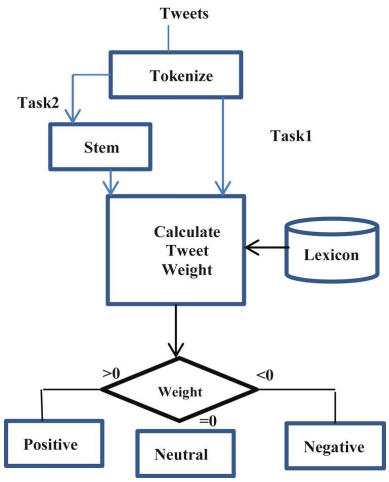
Common Approaches

- Lexicon Based
  - Count based Techniques
  - Rule based Techniques
- Machine Learning/ Statistics Based
  - Naïve Bayes
  - Neural Networks (Deep Learning)
  - And More

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Sentiment Analysis – Lexicon Based



Many open source Lexicons are available
e.g. SentiStrength

Source: Journal of Intelligent & Fuzzy Systems, vol. 29, no. 1, pp. 107-117, 2015

Carnegie Mellon
institute for SOFTWARE RESEARCH

Sentiment Analysis –Naïve Bayes

Assuming parts-of-speech (POS) Tags and n-grams are conditionally independent
M = Twitter message, s = sentiment, G = n-grams, T = POS tags

$$\mathbf{P}(s|M) = \frac{\mathbf{P}(s) * \mathbf{P}(M|s)}{\mathbf{P}(M)}$$

$$\mathbf{P}(s|M) \propto \mathbf{P}(G|s) * \mathbf{P}(T|s)$$

$$\mathbf{P}(s|M) \propto \prod \mathbf{P}(g|s) * \prod \mathbf{P}(t|s) \text{ where } g \in G \text{ and } t \in T$$

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Sentiment Analysis – Naïve Bayes Example

Let's find the sentiment in text "SI is good" using Naïve Bayes.
Assume: + implies Positive sentiment and - implies Negative sentiment, and P(-) = P (+) = 0.5

$$\mathbf{P}(s|M) = \frac{\mathbf{P}(s) * \mathbf{P}(M|s)}{\mathbf{P}(M)}$$

P(+|"SI is good") = P(+)* P("SI is good" | +) / P("SI is good")

P(-|"SI is good") = P(-)* P("SI is good" | -) / P("SI is good")

Divide the last two equations to find the ratio of sentiment:
P(+|"SI is good") / P(-|"SI is good") = P("SI is good" | +) / P("SI is good" | -)

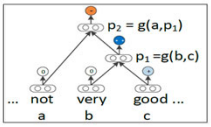
Using independence assumption:
= P("SI"| +) P("good"| +) / P("SI"| -) P("good"| -)
= P("good"| +) / P("good"| -) = .01 / 0.00001 = 1000
=> Sentiment is more positive

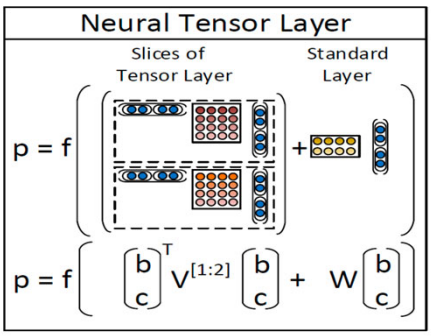
Carnegie Mellon
institute for SOFTWARE RESEARCH

Advanced Sentiment Analysis – Neural Network

Recursive Neural Tensor Network

Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank
Socher et al. 2013





Source: http://cs224d.stanford.edu/lectures/CS224d-Lecture11.pdf

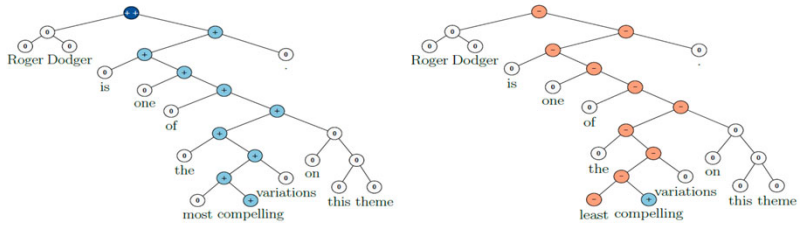
<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Sentiment Analysis – Neural Network

http://nlp.stanford.edu:8080/sentiment/rntnDemo.html

Negation Results



Source: http://cs224d.stanford.edu/lectures/CS224d-Lecture11.pdf

Carnegie Mellon
institute for SOFTWARE RESEARCH

Movie review examples: Stanford Demo:

I think the Academy was too chicken to give the Best Picture to Brokeback Mountain, and that sucks.

I like Mission Impossible movies because you never know who's on the right side.

I loved the Da Vinci Code, but now I want something better and different

Then snuck into Brokeback Mountain, which is the most depressing movie I have ever seen.

I think I hate Harry Potter because it outshines much better reading material out there and the movies are just plain stupid to begin with.

why may you ask well I love Mission Impossible stories.

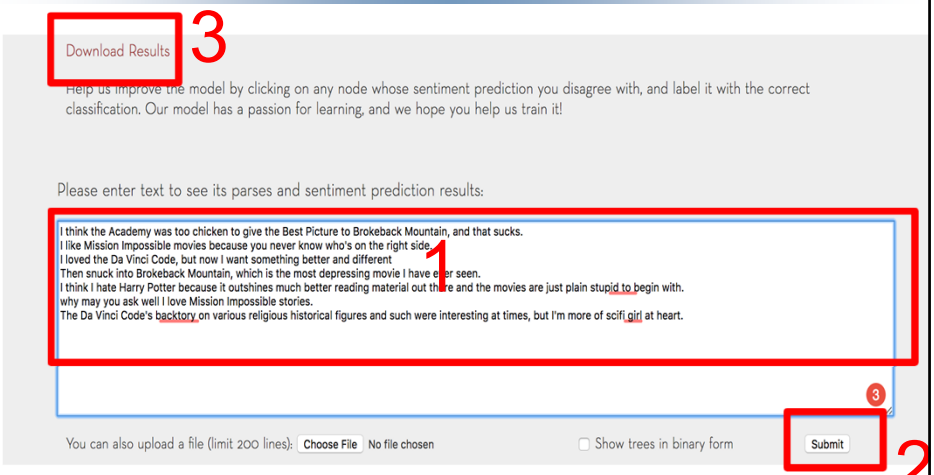
The Da Vinci Code's backtory on various religious historical figures and such were interesting at times, but I'm more of scifi girl at heart.

http://nlp.stanford.edu:8080/sentiment/rntnDemo.html

<Your Name>

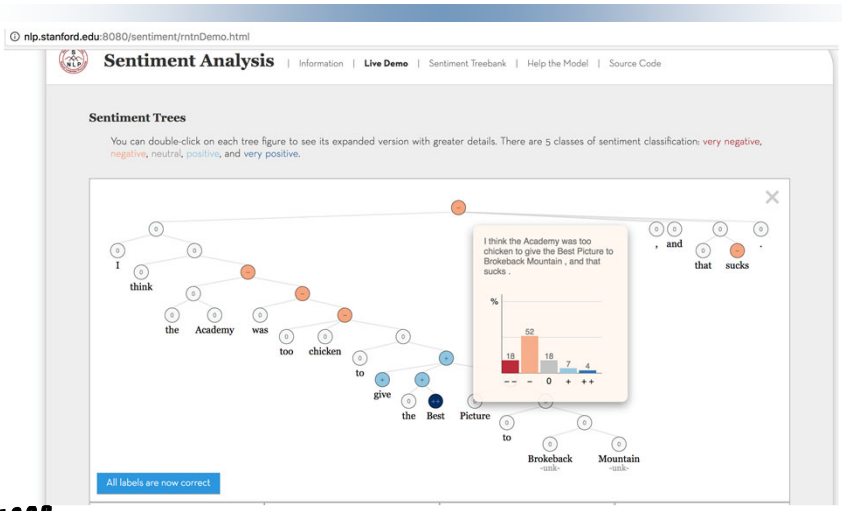
Carnegie Mellon
institute for SOFTWARE RESEARCH

SA using Stanford NN: Input Data



Carnegie Mellon
institute for SOFTWARE RESEARCH

SA using Stanford NN: Input Data



<Your Name>

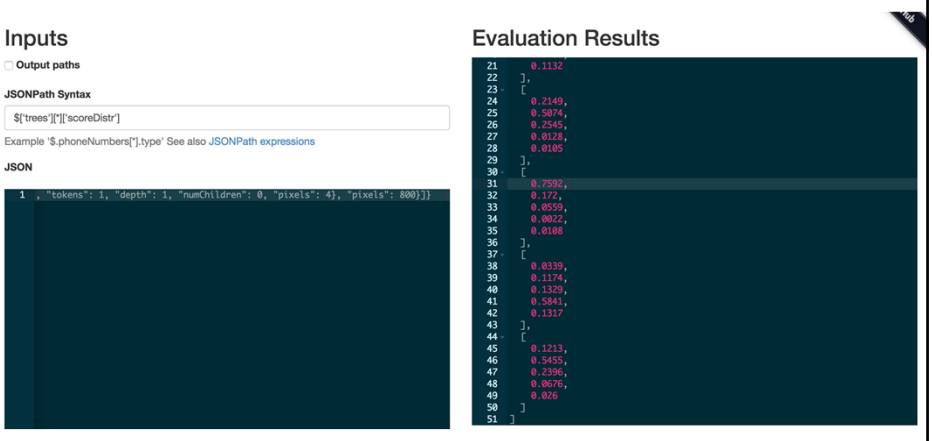
Carnegie Mellon
institute for SOFTWARE RESEARCH

Downloaded Results Need More processing

- Goto:
  - http://jsonpath.com/
  - Use **JSONPath Syntax**
    - $['trees'][*]['scoreDistr']
- Goto:
  - https://konklone.io/json/
  - Convert json to csv
- Final Sentiment Calculation in Excel

Carnegie Mellon
institute for SOFTWARE RESEARCH

Get CSV Output from Stanford Results JSON

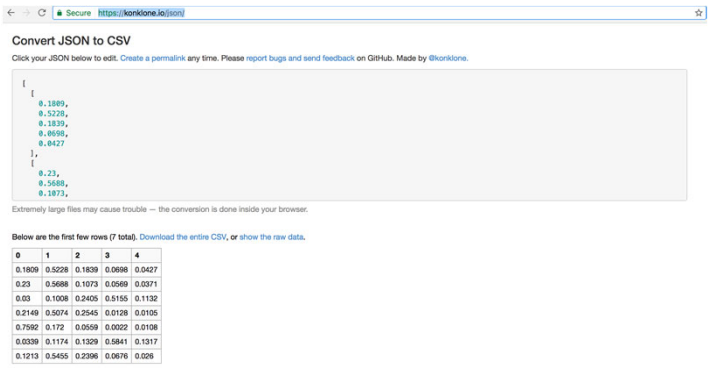


<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Get CSV File and Load in Excel

1. Got to https://konklone.io/json/
2. Paste output of the last page as input
3. Copy output table



Carnegie Mellon
institute for SOFTWARE RESEARCH

SA using Stanford NN: Use Excel to get the final Sentiment Score

- Use Sentiment Calculator Excel sheet
- Paste the result from the last page
- Sheet uses a formula to calculate the final score

Sno	Text	-2	-1	0	1	2	Score
1	"I think the Academy was too chicken to give the Best Picture to Brokeback Mountain , and that sucks .",	0.1809	0.5228	0.1839	0.0698	0.0427	-0.73
2	"I like Mission Impossible movies because you never know who 's on the right side .",	0.23	0.5688	0.1073	0.0569	0.0371	-0.90
3	"I loved the Da Vinci Code , but now I want something better and different",	0.03	0.1008	0.2405	0.5155	0.1132	0.58
4	"Then snuck into Brokeback Mountain , which is the most depressing movie I have ever seen .",	0.2149	0.5074	0.2545	0.0128	0.0105	-0.90
5	"I think I hate Harry Potter because it outshines much better reading material out there and the movies are just plain stupid to begin with .",	0.7592	0.172	0.0559	0.0022	0.0108	-1.67
6	"why may you ask well I love Mission Impossible stories .",	0.0339	0.1174	0.1329	0.5841	0.1317	0.66
7	"The Da Vinci Code 's backtory on various religious historical figures and such were interesting at times , but I 'm more of scifi girl at heart ."	0.1213	0.5455	0.2396	0.0676	0.026	-0.67

<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

More Ways to Get Sentiment

- NetMapper (Demo)
- Stanford Downloadable program (No Demo)
  - https://nlp.stanford.edu/sentiment/code.html

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Example: Fake-News Tweets related to US Politics

Dataset creation steps:

- Collect recent Politifact fake news-headlines related to politics
- Search these headlines on Twitter (daily for a few days)
- Combine the collected tweets as one dataset

<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

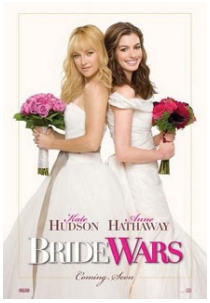
How good are these Sentiment miners?

- How well the result agrees with human judgment?
- Human raters typically agree 79% #
- Recursive Tensor Network in around 85%
- Vader Sentiment Analysis Performance : 96% (F1)
  - May be overfitting

Ogneva, M. "How Companies Can Use Sentiment Analysis to Improve Their Business". Mashable. Retrieved 2012-12-13.

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Anne Hathaway and Warren Buffett have recently been linked in the media—though not romantically, thank god!!



Oct. 3, 2008—Rachel Getting Married opens: BRK.A up .44%

Jan. 5, 2009—Bride Wars opens: BRK.A up 2.61%

Feb. 8, 2010—Valentine's Day opens: BRK.A up 1.01%

March 5, 2010—Alice in Wonderland opens: BRK.A up .74%

Nov. 24, 2010—Love and Other Drugs opens: BRK.A up 1.62%

Nov. 29, 2010—Anne announced as co-host of the Oscars: BRK.A up .25%

Source http://www.cnbc.com/id/42305525

CASOS

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Thank You

Carnegie Mellon
institute for SOFTWARE RESEARCH

Backups

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

SA using Stanford NN: Useful Links

- http://nlp.stanford.edu:8080/sentiment/rntnDemo.html
- http://www.jsoneditoronline.org/
- http://www.convertcsv.com/json-to-csv.htm
- http://sentistrength.wlv.ac.uk/
- http://boston.lti.cs.cmu.edu/classes/95-865-K/HW/HW3/ (Sentiment datasets)
- http://jsonpath.com/?
  - $['trees'][*]['text]
  - $['trees'][*]['scoreDistr']
- https://konklone.io/json/

Carnegie Mellon
institute for SOFTWARE RESEARCH

Can We Build Networks from Text Sentiment?

CASOS