



Sentiment Mining

Sumeet Kumar
sumeetku@cs.cmu.edu

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

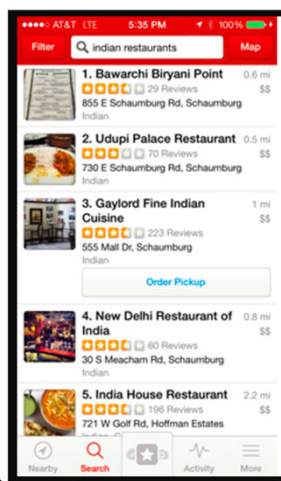


Carnegie Mellon

Center for Computational Analysis of
Social and Organizational Systems
<http://www.casos.cs.cmu.edu/>



Why Sentiment Analysis?



Customer Reviews

JL421 Badonkadonk Land Cruiser/Tank

253 Reviews
5 star: (102)
4 star: (57)
3 star: (29)
2 star: (23)
1 star: (42)

Average Customer Review

★★★★☆ (253 customer reviews)

Share your thoughts with other customers

Create your own review

The most helpful favorable review

2,184 of 2,283 people found the following review helpful:

★★★★★ Finally, a tank you can trust

I'll admit it. Shopping for a personal tank can be a bit daunting. Many times in the past I've purchased overpriced, so-called "battle tanks", then driven them into battle only to be wrecked in ten minutes by the first blow off of some insurgents home-made mortar.

But not this baby, no way.

This tank R-O-C-K-S! Literally- the 400-watt...

[Read the full review >](#)

Published on December 1, 2005 by Thomas Dunham

> See more [5 star](#), [4 star](#) reviews



June 2019

© 2019 CASOS. Director Kathleen M. Carley

2



Carnegie Mellon
IST Institute for Software Research

Why Sentiment Analysis?

Amazon Echo - Black
\$100 online, \$180 nearby ★★★★★ 5,137 product reviews

Reviews

5 star
4 star
3 star
2 star
1 star

4.6
★★★★★
5,137 reviews

ease of use
sound/audio
value
size
design/style
customer service
battery

"Overall the Amazon Echo was easy to set up and use."
"Sound quality amazing!"
"Excellent sound and great value."
"The speaker is also impressive for its size."
"Love it amazing easy to use love the look sound amazing"
"Great customer service"
"Awesome plus made it better by buying the battery!"

★★★★★ **LOVE IT!** – April 3, 2017
Imgiac – Review provided by Bed Bath & Beyond
April 3, 2017
Awesome sound! You must have Amazon Prime at least to get the benefit of the music which is my main interest. I have Prime so it paid for itself already!

★★★★★ **Echo rocks!!** – December 15, 2016
Techgirl0 – Review provided by Bed Bath & Beyond
December 15, 2016
I love Alexa and she is becoming my new BFF! I love it so much that I bought all the men in my life the gift of Alexa for Christmas. Just beginning to scratch the surface with all she can do and improve in my world. Don't wait, get yours now while they are on sales!

★★★★★ **Amazon Echo** – November 21, 2016
NickP – Review provided by Bed Bath & Beyond
November 21, 2016
We have owned the Echo since it was born. Each day we find another use for the Echo. Looked at originally, as a novelty, it fast became a very useful household tool. Our Granddaughters enjoy it so much. We purchased this one for a Christmas Gift. Extremely useful!

CASOS
June 2019 © 2019 CASOS, Director Kathleen M. Carley 3

Carnegie Mellon
IST 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)

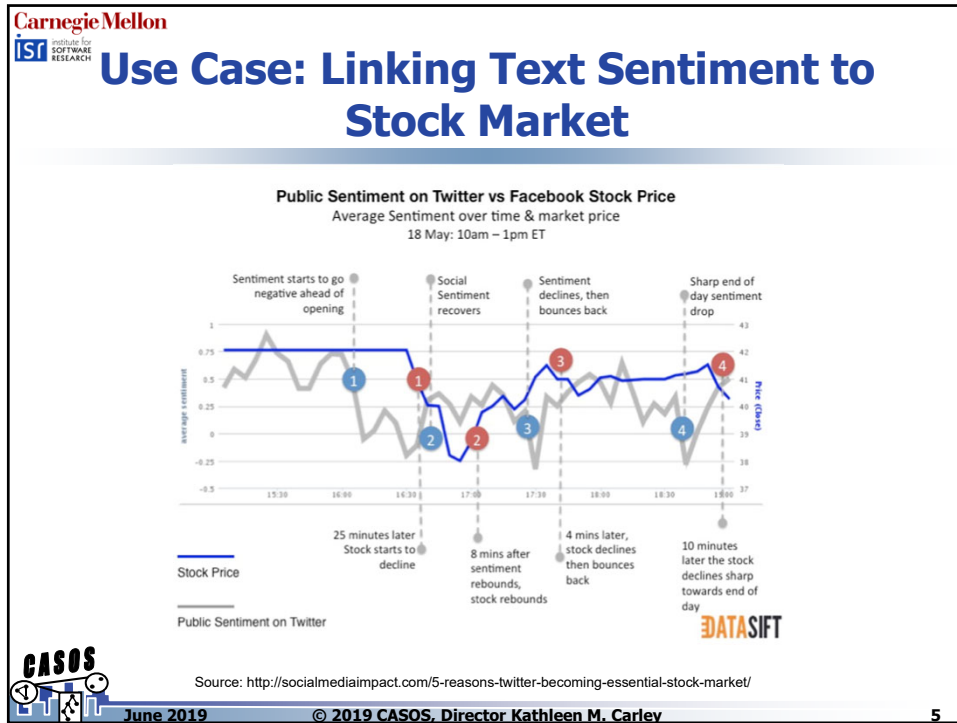
Use Cases:

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

CASOS
June 2019 © 2019 CASOS, Director Kathleen M. Carley 4





Carnegie Mellon
ISI 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

CASOS June 2019 © 2019 CASOS, Director Kathleen M. Carley 6



Carnegie Mellon
ISI Institute for Software Research

Sentiment Analysis – Lexicon Based

```
graph TD; Tweets --> Tokenize; Tokenize -- Task2 --> Stem; Tokenize -- Task1 --> Calculate[Calculate Tweet Weight]; Lexicon[(Lexicon)] --> Calculate; Calculate --> Weight{Weight}; Weight -- ">0" --> Positive[Positive]; Weight -- "=0" --> Neutral[Neutral]; Weight -- "<0" --> Negative[Negative];
```

Many open source Lexicons are available
e.g. SentiStrength

CASOS
June 2019

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

© 2019 CASOS, Director Kathleen M. Carley

7

Carnegie Mellon
ISI 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

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

CASOS
June 2019

© 2019 CASOS, Director Kathleen M. Carley

8



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 $P(-) = P(+) = 0.5$

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

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

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

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:

$$\begin{aligned} &= P("SI" | +) P("good" | +) / P("SI" | -) P("good" | -) \\ &= P("good" | +) / P("good" | -) = .01 / 0.00001 = 1000 \\ &\Rightarrow \text{Sentiment is more positive} \end{aligned}$$



June 2019

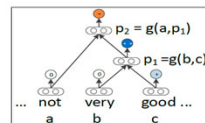
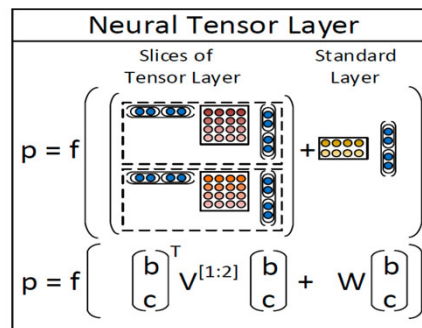
© 2019 CASOS, Director Kathleen M. Carley

9

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>

June 2019

© 2019 CASOS, Director Kathleen M. Carley

10



Carnegie Mellon
ISI 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>

CASOS
June 2019 © 2019 CASOS, Director Kathleen M. Carley 11

Carnegie Mellon
ISI 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 backstory 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>

CASOS
June 2019 © 2019 CASOS, Director Kathleen M. Carley 12

Carnegie Mellon

ISI

Institute for SOFTWARE RESEARCH

SA using Stanford NN: Input Data

Download Results

3

Help us improve the model by clicking on any node whose sentiment prediction you disagree with, and label it with the correct classification. Our model has a passion for learning, and we hope you help us train it!

Please enter text to see its parses and sentiment prediction results:

1

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 backstory on various religious historical figures and such were interesting at times, but I'm more of a sci-fi girl at heart.

3

2

Submit

You can also upload a file (limit 200 lines):

Choose File

No file chosen

☐ Show trees in binary form

CASOS

June 2019

© 2019 CASOS, Director Kathleen M. Carley

13

Carnegie Mellon

ISI

Institute for SOFTWARE RESEARCH

SA using Stanford NN: Input Data

nlp.stanford.edu:8080/sentiment/rntrDemo.html

Sentiment Analysis

Information

Live Demo

Sentiment Treebank

Help the Model

Source Code

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.

1

think

the

Academy

was

too

chicken

to

give

the

Best

Picture

to

Brokeback

Mountain

and

that

sucks

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200


All labels are now correct

CASOS

June 2019


© 2019 CASOS, Director Kathleen M. Carley

14




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



June 2019 © 2019 CASOS, Director Kathleen M. Carley 15



Get CSV Output from Stanford Results JSON

Inputs

☐ Output paths

JSONPath Syntax

`$['trees']['*']['scoreDistr']`


Example '\$.phoneNumbers[*].type' See also JSONPath expressions

JSON

```
1 { "tokens": 1, "depth": 1, "numChildren": 0, "pixels": 41, "pixels": 88011 }
```

Evaluation Results

```
21 0.1132
22 ]
23 [
24 0.2149,
25 0.5874,
26 0.2545,
27 0.0128,
28 0.0185
29 ]
30 [
31 0.7592,
32 0.172,
33 0.8559,
34 0.0022,
35 0.0188
36 ]
37 [
38 0.8339,
39 0.2174,
40 0.1329,
41 0.5841,
42 0.1317
43 ]
44 [
45 0.1213,
46 0.5455,
47 0.2396,
48 0.8676,
49 0.026
50 ]
51 ]
```



June 2019 © 2019 CASOS, Director Kathleen M. Carley 16



Carnegie Mellon

ISI

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

Secure https://konklone.io/json/

Convert JSON to CSV


Click your JSON below to edit. Create a permalink any time. Please report bugs and send feedback on GitHub. Made by @konklone.

```
[
  [
    0.1809,
    0.5228,
    0.1839,
    0.0698,
    0.0427
  ],
  [
    0.23,
    0.5688,
    0.1073,
    0.0569,
    0.0371
  ],
  [
    0.03,
    0.1008,
    0.2405,
    0.5155,
    0.1132
  ],
  [
    0.2149,
    0.5074,
    0.2545,
    0.0128,
    0.0105
  ],
  [
    0.7592,
    0.172,
    0.0559,
    0.0022,
    0.0108
  ],
  [
    0.0339,
    0.1174,
    0.1329,
    0.5841,
    0.1317
  ],
  [
    0.1213,
    0.5455,
    0.2396,
    0.0676,
    0.026
  ]
]
```

Extremely large files may cause trouble — the conversion is done inside your browser.

Below are the first few rows (7 total). Download the entire CSV, or show the raw data.

0	1	2	3	4
0.1809	0.5228	0.1839	0.0698	0.0427
0.23	0.5688	0.1073	0.0569	0.0371
0.03	0.1008	0.2405	0.5155	0.1132
0.2149	0.5074	0.2545	0.0128	0.0105
0.7592	0.172	0.0559	0.0022	0.0108
0.0339	0.1174	0.1329	0.5841	0.1317
0.1213	0.5455	0.2396	0.0676	0.026



June 2019

© 2019 CASOS, Director Kathleen M. Carley

17

Carnegie Mellon

ISI

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	2Score	
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 a scifi girl at heart .",	0.1213	0.5455	0.2396	0.0676	0.026	-0.67





June 2019

© 2019 CASOS, Director Kathleen M. Carley



18





More Ways to Get Sentiment

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



June 2019© 2019 CASOS, Director Kathleen M. Carley19



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




June 2019© 2019 CASOS, Director Kathleen M. Carley20



Carnegie Mellon
ISI Institute for Software Research

How good are these Sentiment miners?

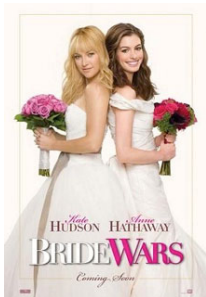
- How well the result agrees with human judgment?
- Human raters typically agree 79% #
- Recursive Tensor Network in around 85%
- Vader Sentiment Analysis Performance : 96% (F1)
 - May be overfitting

 Ognjeva, M. "[How Companies Can Use Sentiment Analysis to Improve Their Business](#)".
Unpublished. Retrieved 2012-12-13.

June 2019 © 2019 CASOS, Director Kathleen M. Carley 21

Carnegie Mellon
ISI 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.cnn.com/id/42305525>



 June 2019 © 2019 CASOS, Director Kathleen M. Carley 22



<Your Name>



Thank You





June 2019



© 2019 CASOS, Director Kathleen M. Carley

23

23



Backups




June 2019

© 2019 CASOS, Director Kathleen M. Carley

24


24





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/>




June 2019

© 2019 CASOS, Director Kathleen M. Carley

25




Can We Build Networks from Text Sentiment?



Twitter post from Climasphe (@climasphe) 3h: The room is filling at @NYUStern to hear @antonioгутerres talk climate action! Tune in live at 4pm: bit.ly/2qyZaD #UNClimateNYU

Yasemin Erboy Ruff (@YErboyRuff) 3h: Looking forward to hearing @UN SG @antonioгутerres speak at 4PM EDT today. Follow via #UNClimateNYU or watch here: youtube.com/unitednations



June 2019

© 2019 CASOS, Director Kathleen M. Carley

26

