

What is sentiment analysis?

How does it work

<http://sentdex.com>

1. "Bag of Words" Model:

- Order doesn't matter
- Doc Classification and Comp Vision
- [Zellig Harris](#)'s 1954 article on *Distributional Structure*
- Email Filtering
- Counts not really shows the importance such as the, ,a, an, to

Example for Bag of Words

(1) John likes to watch movies. Mary likes movies too.

(2) John also likes to watch football games.

Based on these two text documents, a list is constructed as follows:

```
[  
    "John", "likes", "to", "watch", "movies", "also", "football", "games", "Mary", "too"  
]
```

(1) [1, 2, 1, 1, 2, 0, 0, 0, 1, 1]

(2) [1, 1, 1, 1, 0, 1, 1, 1, 0, 0]

Named Entity Recognition

2. Using Natural Language Processing, and the attempt to truly "understand" the text:

- Extracting information, tagging
- Devising models to deal with linguistically complex contexts
- <http://stanfordnlp.github.io/CoreNLP/images/Xi-Jinping.png>

The best way to do sentiment analysis is to use them both effectively!

Why sentiment analysis difficult

- infinite grammatical variations
- misspellings
- slang
- Contextual information hard for machines to interpret

“I like that Corvette.”

“That looks like a Corvette.”

The structure of a solid sentiment test

- data source
- cleanliness of language
- how it is scored
- subject matter
- volume of data tested

How to define accuracy of sentiment

Precision: sentiment rating correct?

Recall: sentiment rated as having tonality, or not having tonality (being neutral)?

F-1 Score: Combination of precision and recall. The score is in a range of 0.0 - 1.0, where 1.0 would be perfect.

How to spot 'bad' sentiment analysis

Run your analysis on a subject that you already know to be widely viewed as negative. If your system classifies as neutral more than 90% of the data, something is probably off.

Why is Sentiment Analysis useful?

- Determine marketing strategy
- Improve campaign success
- Improve product messaging
- Improve customer service
- Test business KPIs
- Generate leads

Example: Environmental non-profit trying to figure out where they should where they should canvas.

Example using sentdex: News outlet deciding on what global topics they should report on (and how they should approach the topic)

Why is sentiment analysis useful in this case?

- Investors
- Sentiment is often most indicative of price in the short term - good for active traders.
- Politicians can use sentiment analysis to see how their actions and choices affect their image in the media.
- Can help you get out of your bubble.

Will Sentiment Analysis ever be 100% accurate, or close?

Probably not.

It's rare that $> 80\%$ of people can agree on the sentiment of a sentence, so machines probably won't do better than that.