

NEW MEDIA AND SENTIMENT MINING

Ravi Malhotra

Ravi.Malhotra@nus.edu.sg

Agenda

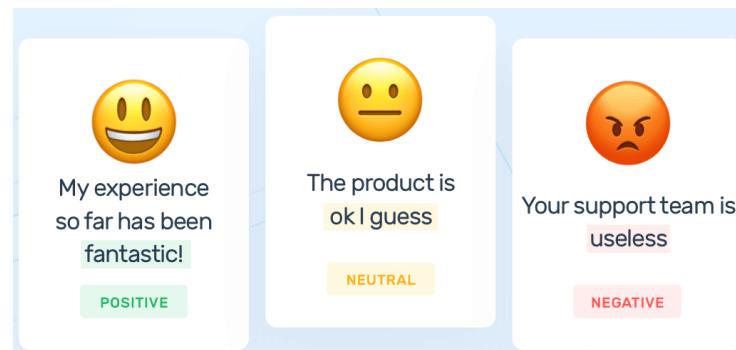
Topics

- Overview of Sentiment Mining
 - Applications of Sentiment Mining
 - Difference from Normal Text Mining/Processing
 - Why it is hard, Some Issues with New media
- Types of Sentiment Mining
 - Introduction to language structure
 - NLP task in Sentiment analysis
 - Type of sentiment analysis & Approaches
 - Tools available
 - Evolution of NLP Models
- Sentiment analysis in Financial & New digital customer era
 - How it effect ? Understand financial phenomenon, sentiment & behavioral fiancé
 - Use of entity & aspect in financial applications
- Class Demo
 - Rules based methods
 - Intro to Spacy & Stanford Stanza
 - How to handle emoji's & emoticons & some advance ways to handle Gifs
 - How to visualize NLP results

Objectives

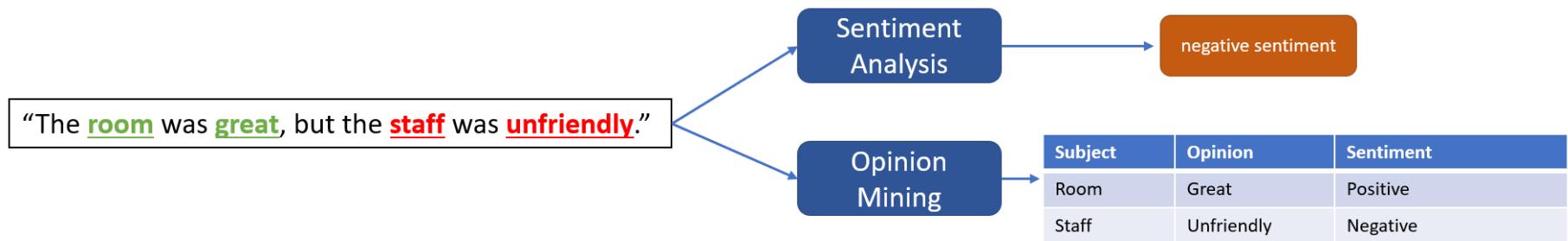
- At the end of this module, you should be able to:
 - Understand the application of sentiment mining
 - Task involved in sentiment mining
 - Type of sentiment mining
 - Tools & methods involved & evolution over the time
 - Sentiment mining use in finance & digital era
 - Visual representation of sentiments & emotions
 - Demo of some essential libraries/tools/methods

Overview of Sentiment Mining



Definition

- Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.
- Generally speaking, sentiment mining seeks to obtain attitude of a **speaker, writer, or other subject with respect to some topic** or the **overall contextual polarity** or **emotional reaction** to a document.



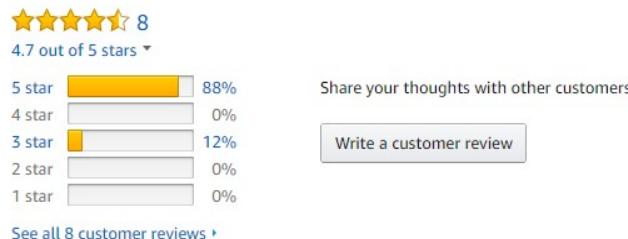
Applications of sentiment mining

- Customer feedback and reviews
- Financial prediction
- Fraud detection of corporate misconduct
- Political ‘sentiment’
- Public Policy

1. Customers' reviews (eg Amazon)

- Social media monitoring, Brand monitoring, Voice of customer (VoC), Customer service

Customer reviews



Top customer reviews

 Lucas N. Santos

★★★★☆ Good overview on current topics

January 15, 2009

Format: Hardcover | [Verified Purchase](#)

What I liked most about the book was the scratch I got when facing all the possibilities regarding data that is free available on the Internet. My interest area is crawling, and there is an exclusive chapter about it on the book. But as with all others chapters, it's only a bird's-eye view on the topic, so specifically the crawler part of the book wasn't of much use. In spite of it, my expectations were reached with the rest of the work, since I just wanted to be aware of what is happening today concerning Web data mining. I must note that, although chapters on relevant topics are small (more or less 30 and so pages) and surely don't cover all the nuances, the book comes with plenty of references for anyone who wants to dig further.

15 people found this helpful

[Helpful](#) [Not Helpful](#) | [Comment](#) | [Report abuse](#)

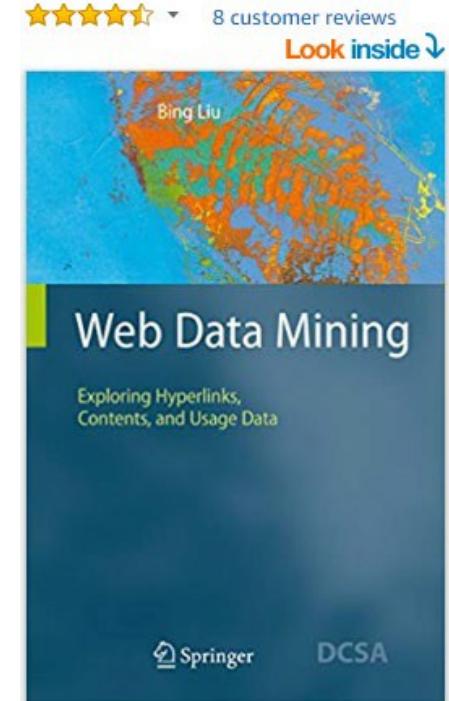
 LOV

★★★★★ A well-written book covering breadth and depth of web data mining

February 9, 2015

Format: Hardcover | [Verified Purchase](#)

This is a very well-written book. It is also a highly ambitious project as the book covers breadth and depth of many web analytics and data mining/machine learning related topics. It is also written in a very accessible way but still delivers strong technical knowledge for technical audience. I don't think this book has much competition in this area (so far), and the author clearly is a real expert.



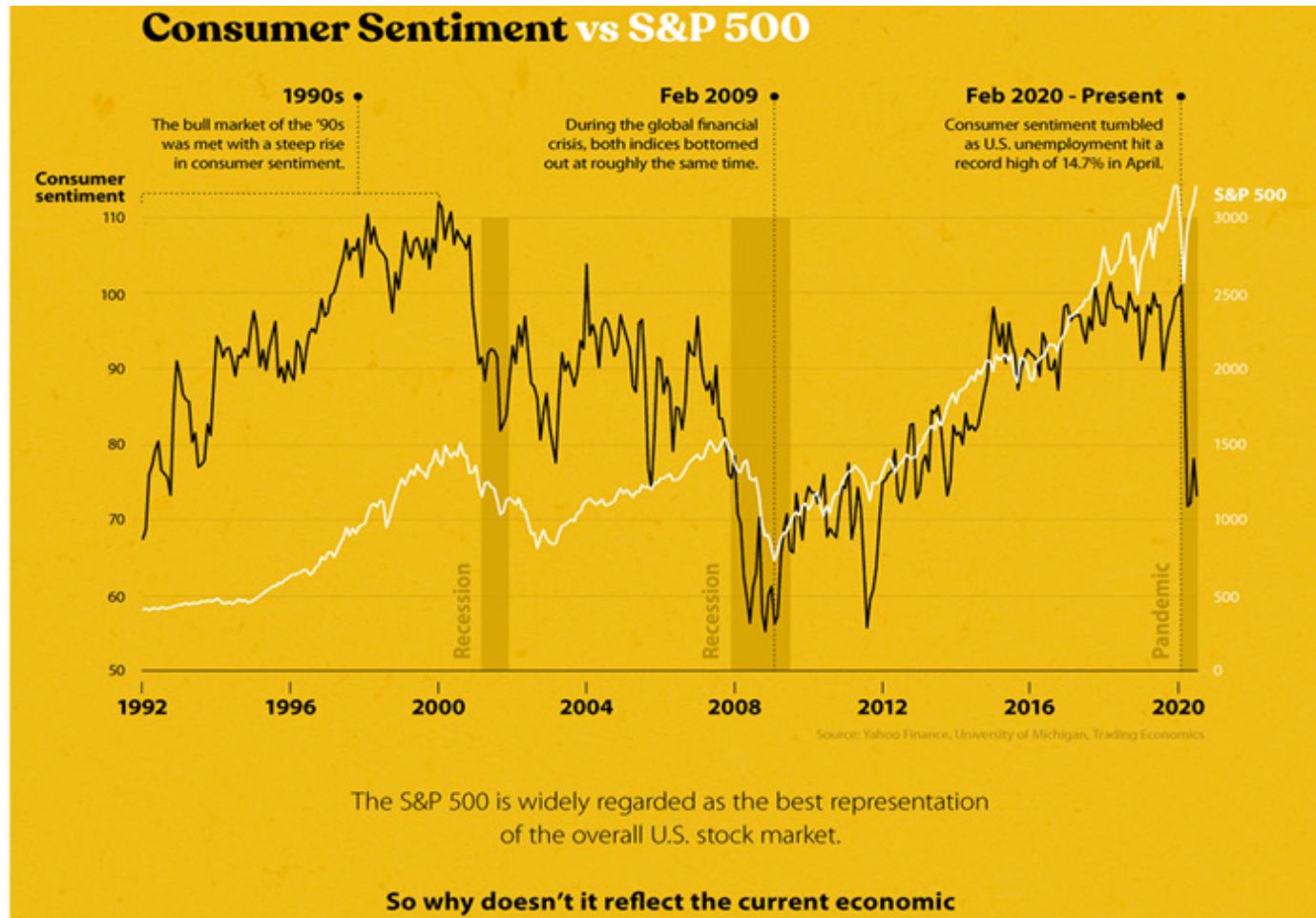
ISBN-13: 978-3540378815

ISBN-10: 3540378812

[Why is ISBN important?](#) ▾

2. Financial prediction

- Sentiment drives the stock market. It changes people's consumption habits which in turn drives their investment and savings tendencies.



3. Corporate misconduct

- Early warning of fraud and misconduct in chat messaging and emails.

Finance in Focus: Meet Watson—your new surveillance officer

NOVEMBER 11, 2016 | 12:29



Questionable behavior and employee misconduct have cost financial institutions more than \$200 billion in fines and settlements since 2009. The extreme cost of this misconduct is proving to be a daunting challenge for financial services institutions. Although surveillance offers a means of protecting investors, consumers and firms from fraud and misconduct, current surveillance methods are falling short. Accordingly, financial institutions are searching for new ways of addressing this ever-present problem.

Moving surveillance out of the silo

Traditional surveillance systems are not working, in part because market manipulation techniques are constantly changing. Even



<https://www.ibmbigdatahub.com/podcast/finance-focus-meet-watson- your-new-surveillance-officer>

Enron emails corpus

- Good ex: is the Enron corporate scandal.
- The open corpus (in Carnegie Mellon Univ.) has given rise to whole research study in NLP.
- Most of these research concludes that corporate scandals can be predicted through email correspondence ~
- Since then, it has resulted in a corporate culture shift in how companies communicate.
- Data set link if you want to play around:
- <https://www.cs.cmu.edu/~enron/>
- <https://www.ceas.cc/papers-2004/168.pdf>
- Researchers were able to find that during the crisis period, **communication** among employees became **more diverse** & changing sentiment concerning established contacts and formal roles. In the crisis period, previously disconnected employees began to engage in mutual communication, so that interpersonal communication intensified and spread through the network, bypassing formal communication chains.

4. Political sentiment

- US Presidential campaign – Trump vs Clinton

PACIS 2017 PROCEEDINGS

How Trump won: The Role of Social Media Sentiment in Political Elections

[Chong Oh](#), University of Utah

[Follow](#)

[Savan Kumar](#), University of Utah

[Follow](#)

Abstract

The outcome of the recent US Presidential Election of 2016 shocked and baffled many. Some claimed that social media may play a larger role in influencing the outcome than expected. This study examined Twitter messages containing political discussions with references to both Trump and Clinton to uncover insights about the role of social media sentiment in political elections. We adhere to the social media analytics (SMA) framework of Fan and Gordon (2014) and the sentiment analysis taxonomy of Abbasi, Chen, and Salem (2008) as a structure to extract positive and negative sentiment from the collected tweets during the pre-election period between Nov 3 and Nov 7. The first finding reveals that Trump has an overwhelmingly larger volumes of total, positive, and negative tweets over Clinton implying a higher volume of public discourse around Trump. Secondly, the propagation of negativism towards Clinton is much more than Trump although both candidates have increasingly more negative tweets days leading up to the Election Day of Nov 8. Finally, word clouds for both candidates reveal that the Twitter public are engrossed with more negative topics against Clinton than Trump. This study clarifies the role of social media sentiment, specifically in how Trump

 [Download](#)

235 DOWNLOADS

Since September 11, 2017

 [PLUMX METRICS](#)

SHARE



4. Political sentiment (cont'd)

- Russia's role in social media in Trump campaign using bots, fake Twitter accounts & FB advertisements
- Uses IRA (Internet Research Agency), friends of Russian Intelligence to increase clicks on target.t

≡ FORTUNE

LEADERSHIP • DONALD TRUMP

How Russians Used Social Media to Boost the Trump Campaign, According to Robert Mueller's Indictment



5. Public Policy

- Identify Fraud
- Understand public sentiments about **unmet needs**
- Find **emerging concerns** that can shape policy
- Active participation while policy making
- **Geographical split** of emotions and sentiments
- **Temporal shift** in sentiments or due to events
- This can also link back to business and financial market performance.

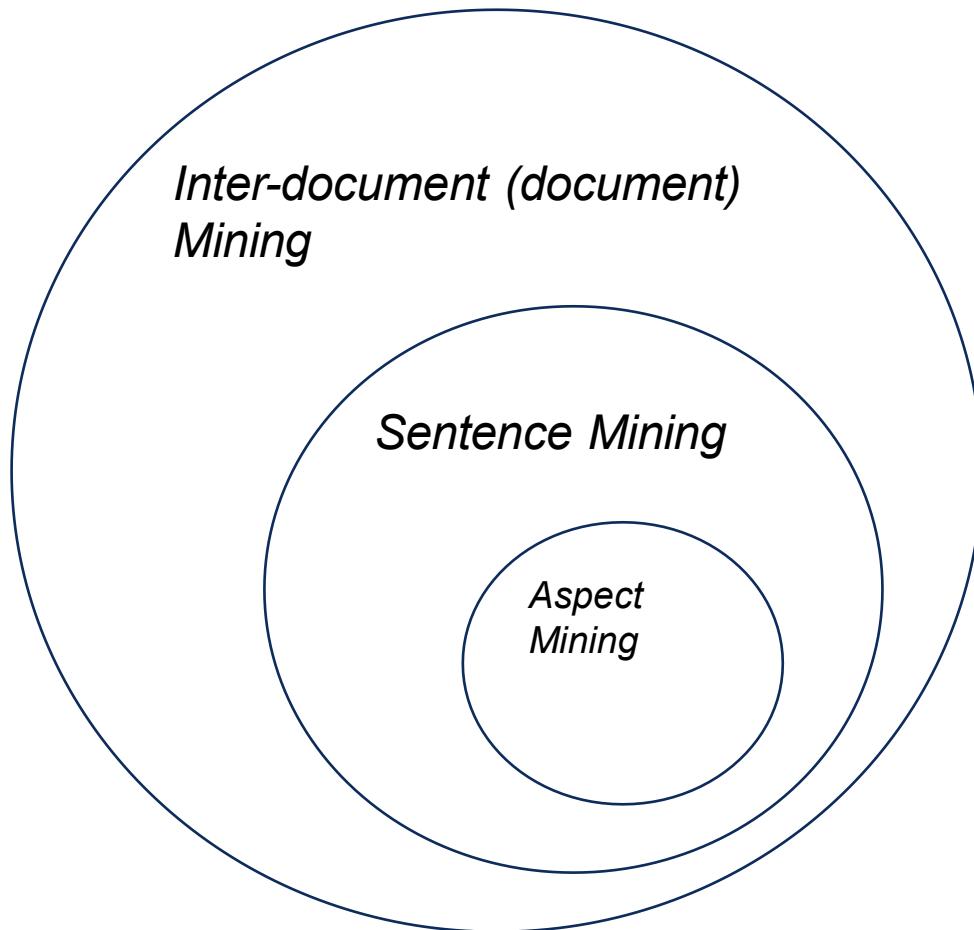
Sentiment & Text Mining

- Text mining is a subset of data mining and is defined as the transformation of **unstructured data** into **answers** to **business questions**
- Sentiment mining largely understood as a classification technique presently.
 - Not just positive/ negative polarity though;
 - Can be expanded to include other text classification tasks.
- Pipeline for text mining and sentiment mining task can be very similar but for sentiment mining it will be very sensitive to words meaning (**if we are detecting emotion**), usage (**Sentiment & emotion both**) and very sentiment aware (**can handle negations, sarcasm & emoji's etc.**).

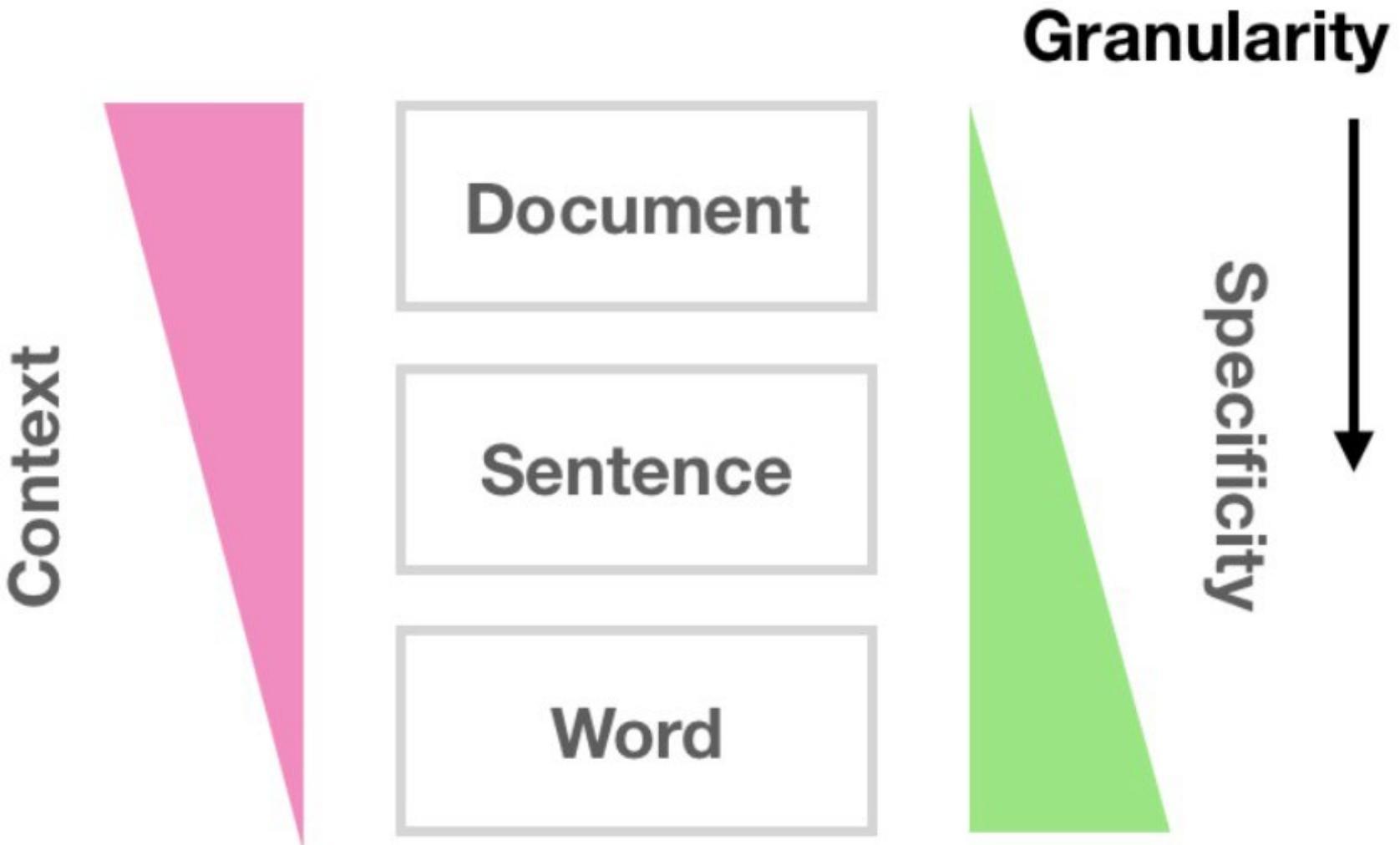
Core task of sentiment mining

- To be able to extract quadruple (**o, t, t, s**)
- These are:
 - **Opinion holder:** who expressed the opinion (o)
 - **Opinion target:** Entity + (optionally) aspect about which the opinion was expressed (t)
 - **Score:** +ve or -ve, or some scale & granularity (s)
 - **Time:** when the opinion was expressed (t)
- The time aspect (t) and opinion holder (o) are important when it comes to sentiment aggregation.

Main levels of sentiment mining



Main levels of sentiment mining



One loses context at higher granularity and specificity in lower granularity.

(Inter-) Document level sentiment mining

- Documents include blogs, articles, news etc. usually longer.
- Key issues include ‘too many items talked about’ and how to ‘link them’ together. Not wise to allocate a single polarity to the whole context of document.
- An approach is to model the general topic, or unearth the individual entities and relationships in document and do granular polarity assignment.

Why is sentiment analysis hard?

- Language nuances
 - **Rational Evaluation** express no emotion
 - Evaluation of aspects/features: how to **quantify** then?
 - E.g., “The camera takes 2 secs to start up”
 - Contrast with “The camera starts up very fast.”
 - **Emotional words** may not express a sentiment
 - E.g., “**Amazingly**, the stewardesses treat is as normal”
 - E.g., “The phone is an **amazing** birthday present”
- Slang! Eg. can...
- Sarcasm/ irony
 - What do you mean?
 - What are you trying to say?
 - Donald Trump’s tweets
- Often a mix of intuition and science to do a good sentiment analysis engine. Every system is bespoke.

Sentiment mining today

- Some domain problems are easier than the rest
 - Eg restaurants, movies and hotels because they are much studied and applied
 - Can you transfer know-how from one domain to another? Not with extreme difficulty!
- In computational linguistics literature, ~88% accuracy in SemEval competitions is very good.

New Media issues

- Availability of meta-data
- Opinion spam
- Fake news
- Influencers
- Virality of posts

1. Availability of meta-data

- New media is about user generated content (UGC) data
- Meta-data includes ‘likes’, ratings (overall sentiment), author, date/ time etc. Use them.

Reviews at Wild Honey

[Add Review !\[\]\(c6747d08ffcbb3c0701a343df825d2f1_img.jpg\)](#)

CHUBBY BOTAK KOALA • 07 Oct 2011 • 406 Reviews

Around the world with Breakfast



This place has been buzzing with review since it opened in late 2009. During the weekend, it is notoriously famous for long waiting time. 76 reviews in HGW and counting, but this does not discourage me in penning my 2 cents. ...

[Read more](#)

Must Tries: English breakfast, Scandinavian breakfast

2. Opinion Spam (Fake) Detection

- Opinion spam refers to **non-genuine** opinion found on websites, reviews, etc.
 - Undeserving positive reviews – promote products
 - Malicious negative reviews – damage reputations
 - In general, “reviews” (positive or negative) where the “opinion holder” did not honestly and actually review the product.
 - E.g., a **positive** review of a **very good** book by someone who had not read the book would be opinion spam
- Opinion spam is a business
 - Pay for positive reviews for own business, negative reviews for competitors

Is this review fake or not?

I want to make this review in order to comment on the excellent service that my mother and I received on the Serenade of the Seas, a cruise line for Royal Caribbean. There was a lot of things to do in the morning and afternoon portion for the 7 days that we were on the ship. We went to 6 different islands and saw some amazing sites! It was definitely worth the effort of planning beforehand. The dinner service was 5 star for sure. One of our main waiters, Muhammad was one of the nicest people I have ever met.

However, I am not one for clubbing, drinking, or gambling, so the nights were pretty slow for me because there was not much else to do. Either than that, I recommend the Serenade to anyone who is looking for excellent service, excellent food, and a week full of amazing day-activities!

[https://www.inc.com/jessica-stillman/heres-how-to-spot-fake-online-reviews-with-90- perc.html](https://www.inc.com/jessica-stillman/heres-how-to-spot-fake-online-reviews-with-90-perc.html)

<https://www.cs.uic.edu/~liub/FBS/fake-reviews.html>

2. Opinion Spam (Fake) Detection

- Most spam reviews detector works for only reviews generated by AI.
- Not quite work if it is written by humans. Hard to tell if he or she actually use the product (unless verified buyer)
- Fake news often sensationalize which can impact people's opinions greatly.

Fake News: Lies spread faster on social media than truth does

People are quicker to repeat something that's wrong than something that's true

by Maggie Fox / Mar.09.2018 / 3:05 AM ET / Updated Mar.09.2018 / 8:51 PM ET

Singapore's proposed fake news law worries social media players

⌚ March 23, 2018 🚩 News Centre 📰 News 💬 0