Data Science Foundations

Master in Big Data Solutions 2017-2018

Liana Napalkova

liana.napalkova@bts.tech

Ludovico Boratto

ludovico.boratto@bts.tech

Francisco Guitierres

francisco.gutierres@bts.tech

Ⓒ Barcelona Technology School S.L.

Session 19
Introduction to Sentiment Analysis

What is Sentiment?

- Sentiment = Feelings
  - Attitudes
  - Emotions
  - Opinions
- Subjective impressions, not facts

What is Sentiment?

- What is Sentiment?
- Generally, a binary opposition in opinions is assumed
- For/against, like/dislike, good/bad, etc.
- Some sentiment analysis jargon:
  - "Semantic orientation"
  - "Polarity"

What is Sentiment Analysis?

- Using NLP, statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit

- Sometimes refered to as opinion mining, although the emphasis in this case is on extraction

Questions SA might ask

- Is this product review positive or negative?

- Is this customer email satisfied or dissatisfied?

- Based on a sample of tweets, how are people responding to this ad campaign/product release/news item?

- How have bloggers' attitudes about the president changed since the election?

Other related tasks

- Information extraction (discarding subjective information)

- Question answering (recognizing opinion-oriented questions)

- Summarization (accounting for multiple viewpoints)

Other related tasks

- "Flame" detection
- Identifying child-suitability of videos based on comments
- Bias identification in news sources
- Identifying (in)appropriate content for ad placement

Applications in Business Intelligence

- Question: "Why aren't consumers buying
- our laptop?"
- We know the concrete data: price, specs, competition, etc.
- We want to know subjective data: "the design is tacky," "customer service was condescending"
- Misperceptions are also important, e.g., "updated drivers aren't available" (even though they are)

Applications in Business Intelligence

- It is very difficult to survey customers who didn't buy the company's laptop

- Instead, you could use SA to

  A. search the web for opinions and reviews of this and competing laptops. Blogs, Epinions, amazon, tweets, etc.

  B. create condensed versions or a digest of consensus points

Cross domain applications

- Insights and applications from SA have been useful in other areas
  - Politics/political science
  - Law/policy making
  - Sociology
  - Psychology

Political SA

- Numerous applications and possibilities
- Analyzing trends, identifying ideological bias, targeting advertising/messages, gauging reactions, etc.
- Evaluation of public/voters' opinions
- Views/discussions of policy

SA and Sociology

- Idea propagation through groups is an important concept in sociology (cf. Rogers 1962, *Diffusion of Innovations*)

- Opinions and reactions to ideas are relevant to adoption of new ideas

- Analyzing sentiment reactions on blogs can give insight to this process

- E.g., Kale et al (2007), *Modeling trust and influence in the blogosphere using link polarity*

SA and Psychology

- Potential to augment psychological investigations/experiments with data extracted from NL text

- Dream sentiment analysis (Nadeau et al., 2006)

In general

- Humans are subjective creatures and opinions are important. Being able to interact with people on that level has many advantages for information systems.

Challenges in SA

- People express opinions in complex ways
- In opinion texts, lexical content alone can
- be misleading
- Intra-textual and sub-sentential reversals,
- negation, topic change common
- Rhetorical devices/modes such as sarcasm, irony, implication, etc.

A letter to a hardware store

Dear <hardware store>

   Yesterday I had occasion to visit <your competitor>. The had an excellent selection, friendly and helpful salespeople, and the lowest prices in town.

   You guys suck.

                              Sincerely,"

What to classify

- There are many possibilities for what we might want to classify:
    - Users
    - Texts
    - Sentences (paragraphs, chunks of text?)
    - Predetermined descriptive phrases (<ADJ N>, <N N>, <ADV ADJ>, etc)
    - Words
    - Tweets/updates

Classifying words/short phrases

- The building blocks of sentiment expression
- Short phrases may be just as important as words:
  - "lowest prices"
  - "high quality"
- We need an approach to deal with these before moving on to other classification tasks

Polarity keywords

- There seems to be *some* relation between positive words and postive reviews

- Can we come up with a set of keywords by hand to identify polarity?

Keyword methods

- Data-driven methods can be used to generate keyword lists that model better than human-generated keyword lists

- Unigram methods on similar data have reached 80% accuracy (Pang et al, 2002)

- Not bad, but lower than you'd usually see in topic-based binary text classification

Smileys

- A common approach for working with tweets and short text updates

- Very little text to work with

- Sentiment most succinctly represented with emoticons/smileys

SOME ACTUAL EXAMPLES OF SENTIMENT TEXT

Amazon (5 star)

- "The characters are so real and handled so carefully, that being trapped inside the Overlook is no longer just a freaky experience. You run along with them, filled with dread, from all the horrible personifications of evil inside the hotel's awful walls. There were several times where I actually dropped the book and was too scared to pick it back up. Intellectually, you know it's not real. It's just a bunch of letters and words grouped together on pages. Still, whenever I go into the bathroom late at night, I have to pull back the shower curtain just to make sure."

Amazon.com (1 star)

- "The original Star Wars trilogy was a defining part of my childhood. Born as I was in 1971, I was just the right age to fall headlong into this amazing new world Lucas created. I was one of those kids that showed up early at toy stores […] anxiously awaiting each subsequent installment of the series. I'm so glad that by my late 20s, the old thrill had faded, or else I would have been EXTREMELY upset over Episode I: The Phantom Menace… perhaps the biggest let-down in film history."

Pitchfork.com (0.0 out of 10)

- "Ten years on from Exile, Liz has finally managed to achieve what seems to have been her goal ever since the possibility of commercial success first presented itself to her: to release an album that could have just as easily been made by anybody else."

Amazon.com (1 star)

- "It took a couple of goes to get into it, but once the story hooked me, I found it difficult to put the book down -- except for those moments when I had to stop and shriek at my friends, "SPARKLY VAMPIRES!" or "VAMPIRE BASEBALL!" or "WHY IS BELLA SO STUPID?" These moments came increasingly often as I reached the climactic chapters, until I simply reached the point where I had to stop and flail around laughing."