



Basics of Sentiment Analysis

Dr. Goutam Chakraborty



Outline

- Explain why sentiment analysis is important.
- Define sentiment analysis.
- Describe generic challenges in doing sentiment analysis.
 - Different types of sentences in social media
 - Sentiments at different granularity levels
- Describe different types of classification models used in sentiment analysis.



Why Is Sentiment Important?

- *“In this age, in this country, public sentiment is everything. With it, nothing can fail; against it, nothing can succeed. Whoever molds public sentiment goes deeper than he who enacts statutes, or pronounces judicial decisions.”*



Why Do We Need Sentiment Analysis for Business?

- Opinions (sentiments) have always been key influencers of behaviors.
 - Consumers' beliefs, perceptions, and behavior are often influenced by what others think and do.
 - Individuals often ask opinions from friends and family.
 - Organizations use focus groups, opinion polls, and so on, to explicitly ask about brands and company
- Impact of the web and the rise of social media
 - Scale is global, no longer a close circle of friends.
 - Personal experiences and opinions about almost anything are freely expressed by consumers in reviews, forums, blogs, networking sites, Twitter, and so on.
 - Opinions abound in news articles and commentaries.



What's in a Name?

- Sentiment analysis, opinion mining, sentiment mining, subjectivity analysis, affect analysis, emotion detection...
- Whatever you call it, it is an intellectually challenging problem and has many applications.
 - It originated in computer science and NLP and has now spread to management and business.
 - It is difficult to do, but rewarding!



Basics of Sentiment Analysis

- The basic task involved in sentiment analysis is *identification* and *quantification* of *polarity* of sentiments (such as positive, negative, neutral, or mixed) expressed in written opinions, expressions, reviews, and comments.
 - It involves many of the text analytics steps such as tokenization, sentence identification, and POS tagging that have been discussed before.
 - Sentiment analysis has to go beyond the basic steps mentioned in text analytics, mostly due to the nature of free-flowing text typically found in social media.



Types of Sentences Often Found in Opinions Expressed in Social Media

- Declarative
- Imperative
- Interrogative
- Comparative
- Non-comparative



An Example Using a TV Review

The TV is wonderful. Great size, great picture, easy interface. It makes a cute little song when you boot it up and when you shut it off. I just want to point out that the 43" does not in fact play videos from the USB. This is really annoying because that was one of the major perks I wanted from a new TV. Looking at the product description now, I realize that the feature list applies to the X758 series as a whole, and that each model's capabilities are listed below. Kind of a dumb oversight on my part, but it's equally stupid to put a description that does not apply on the listing for a very specific model.



Granularity Levels for Sentiment Analysis

- Document level
- Sentence level
- Entity (or *product*) and attribute (or *aspect* or *feature*) level
- Hierarchical taxonomy



Challenges in Doing Sentiment Analysis

- Among the challenges are all of the typical NLP problems associated with analyzing text data, such as correctly identifying POS tags, disambiguation of terms and lexicons, and spelling errors.
- Then we have to identify polarity of opinion terms. Some terms have polarity independent of context (such as *beautiful*).
 - But sometimes polarity of opinion terms can change based on context. Consider an example: Size seems small.
 - Is the word *small* positive or negative?
 - Nature of the communication medium
 - Tweets versus posting on blogs and review sites



Two General Classes of Sentiment Classification

- Unsupervised classification
- Supervised classification



Unsupervised Classification

- Typically, this is done at the *sentence level*.
- It uses rules that combine opinion words and their *dependency relationships within a sentence* with products and features obtained through parsing.
- Example: The picture quality is outstanding.



Supervised Classification

- Typically, this is done at the *document level*.
- Depending on data availability, any commonly used predictive models may be used.
- The primary challenge for modelers is to select relevant input variables from text features such as *terms and their frequencies and POS tags*.