

Sentiment Analysis

Ayoub Bagheri & Anastasia Giachanou

The Little Prince example

This is a nice book for both young and old. It gives beautiful life lessons in a fun way. Definitely worth the money!

+ Educational

+ Funny

+ Price

Nice story for older children.

+ Funny

- Readability

ANTOINE DE SAINT-EXUPÉRY

The Little Prince

The original
translation by
Katherine Woods,
with full-colour
illustrations

Sentiment

- Sentiment =
 - Feelings, Attitudes, Emotions, Opinions
 - A thought, view, or attitude, especially one based mainly on emotion instead of reason
- Subjective impressions, not facts

Webster's dictionary

Definition of *sentiment*

- 1 **a** : an attitude, thought, or judgment prompted by feeling : [PREDILECTION](#)
b : a specific view or notion : [OPINION](#)
- 2 **a** : [EMOTION](#)
b : refined feeling : delicate sensibility especially as expressed in a work of art
c : emotional idealism
d : a romantic or nostalgic feeling verging on [sentimentality](#)
- 3 **a** : an idea colored by emotion
b : the emotional significance of a passage or expression as distinguished from its verbal context

Webster's dictionary

Definition of *opinion*

- 1 **a** : a view, judgment, or appraisal formed in the mind about a particular matter
// We asked them for their *opinions* about the new stadium.
b : [APPROVAL](#), [ESTEEM](#)
// I have no great *opinion* of his work.
- 2 **a** : belief stronger than impression and less strong than positive knowledge
// a person of rigid *opinions*
b : a generally held view
// news programs that shape public *opinion*
- 3 **a** : a formal expression of judgment or advice by an expert
// My doctor says that I need an operation, but I'm going to get a second *opinion*.
b : the formal expression (as by a judge, court, or referee) of the legal reasons and principles upon which a legal decision is based
// The article discusses the recent Supreme Court *opinion*.

Scherer typology of affective states

- Emotion: brief organically synchronized ... evaluation of a major event
 - angry, sad, joyful, fearful, ashamed, proud, elated

- Mood: diffuse non-caused low-intensity long-duration change in subjective feeling
 - cheerful, gloomy, irritable, listless, depressed, buoyant
- Interpersonal stances: affective stance toward another person in a specific interaction
 - friendly, flirtatious, distant, cold, warm, supportive, contemptuous
- Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons
 - liking, loving, hating, valuing, desiring
- Personality traits: stable personality dispositions and typical behavior tendencies
 - nervous, anxious, reckless, morose, hostile, jealous

Sentiment analysis

- Use of natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from unstructured text
- Other terms
 - Opinion mining
 - Sentiment mining
 - Subjectivity analysis

Sentiment analysis | can be applied in every topic & domain

- Book: is this review positive or negative?
- Humanities: sentiment analysis for German historic plays.
- Products: what do people think about the new iPhone?
- Blog: how are people thinking about immigrants?
- Politics: who is going to win the election?
- Twitter: what is the trend today?
- Movie: is this review positive or negative (IMDB, Netflix)?
- Marketing: how is consumer confidence? Consumer attitudes?
- Healthcare: are patients happy with the hospital environment?

Opinion types

- Regular opinions: Sentiment/opinion expressions on some target entities
 - Direct opinions:
 - * “The touch screen is really cool.”
 - Indirect opinions:
 - * “After taking the drug, my pain has gone.”
- Comparative opinions: Comparison of more than one entity.
 - E.g., “iPhone is better than Blackberry.”

Practical definition

- An opinion is a quintuple (entity, aspect, sentiment, holder, time) where
 - entity: target entity (or object).
 - Aspect: aspect (or feature) of the entity.
 - Sentiment: +, -, or neu, a rating, or an emotion.
 - holder: opinion holder.
 - time: time when the opinion was expressed.

Sentiment analysis

- Simplest task:
 - Is the attitude of this text positive or negative?
- More complex:
 - Rank the attitude of this text from 1 to 5
- Advanced:
 - Detect the target, source, or complex opinion types
 - Implicit opinions or aspects

Simple task: Opinion summary

Aspect/feature Based Summary of opinions about iPhone:

Aspect: Touch screen Positive: 212

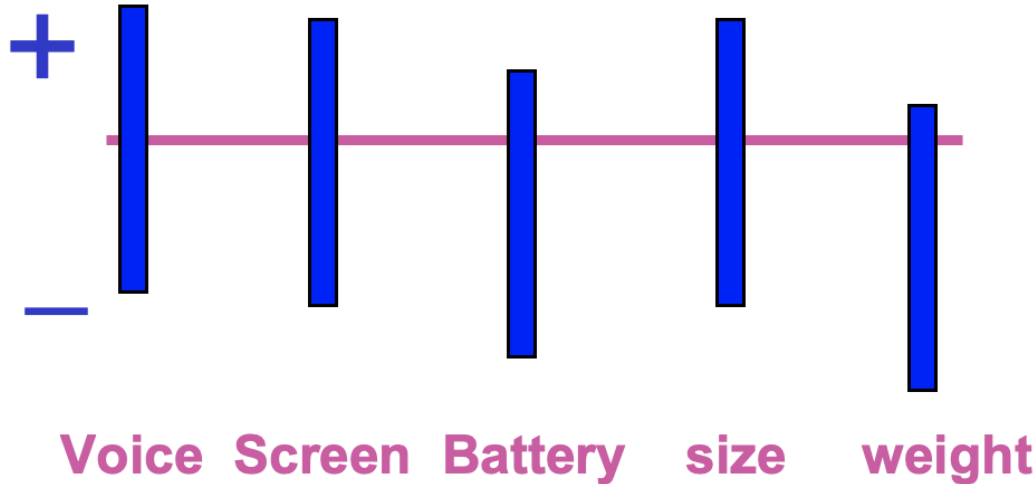
The touch screen was really cool. The touch screen was so easy to use and can do amazing things.

... Negative: 6

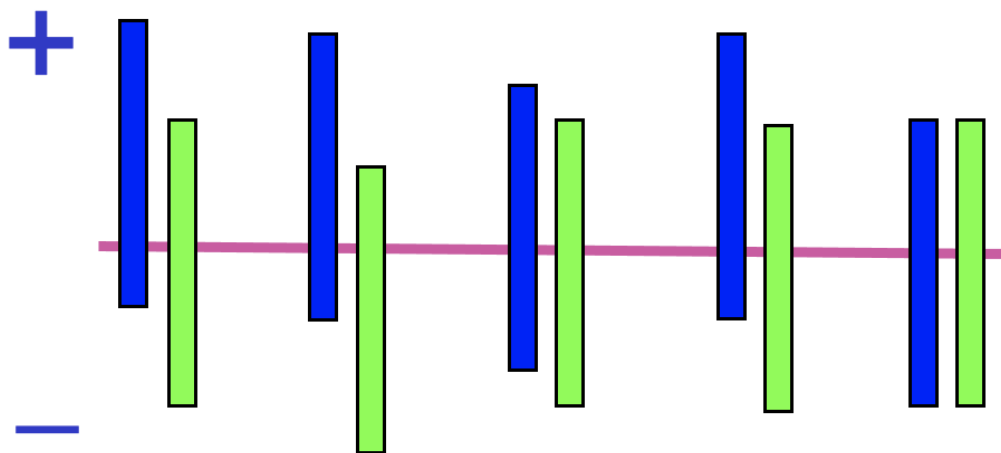
The screen is easily scratched. I have a lot of difficulty in removing finger marks from the touch screen.

... Aspect: Size ...

Opinion Summary of 1 phone



Opinion comparison of 2 phones



Problem

- Which features to use?
 - Words (unigrams)
 - Phrases/n-grams
 - Sentences
- How to interpret features for sentiment detection?

- Bag-of-words
- Annotated lexicons (WordNet, SentiWordNet)
- Syntactic patterns
- Paragraph structure
- Word embedding

Challenges

- Harder than topical classification, with which bag of words features perform well
- Must consider other features due to...
 - Subtlety of sentiment expression
 - * irony
 - * expression of sentiment using neutral words
 - Domain/context dependence
 - * words/phrases can mean different things in different contexts and domains
 - Effect of syntax on semantics

Approaches for sentiment analysis

- **Lexicon-based (dictionary-based) methods**
 - Using sentiment words and phrases: good, wonderful, awesome, troublesome, cost an arm and leg
 - Not completely unsupervised!
- **Supervised learning methods:** to classify reviews into positive and negative.
 - Naïve Bayes
 - Maximum Entropy
 - Support Vector Machine
 - Deep learning

Lexicon-based Methods

LIWC (Linguistic Inquiry and Word Count)

- Home page: <http://liwc.wpengine.com/>
- 2300 words, >70 classes
- Affective Processes
 - negative emotion (bad, weird, hate, problem, tough)
 - positive emotion (love, nice, sweet)
- Cognitive Processes
 - Tentative (maybe, perhaps, guess), Inhibition (block, constraint)
- Pronouns, Negation (no, never), Quantifiers (few, many)

Bing Liu opinion lexicon

- Bing Liu's Page on Opinion Mining
- <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>
- 6786 words
 - 2006 positive
 - 4783 negative

SentiWordNet

- <https://github.com/aesuli/SentiWordNet>
- All WordNet synsets automatically annotated for degrees of positivity, negativity, and neutrality/objectiveness
- [estimable(J,3)] “may be computed or estimated”

Pos 0 Neg 0 Obj 1

- [estimable(J,1)] “deserving of respect or high regard”

Pos .75 Neg 0 Obj .25

Turney algorithm

1. Extract a phrasal lexicon from reviews
2. Learn polarity of each phrase
3. Rate a review by the average polarity of its phrases

<https://arxiv.org/abs/cs/0212032>

Extract two-word phrases with adjectives

First Word	Second Word	Third Word (not extracted)
JJ	NN or NNS	anything
RB, RBR, RBS	JJ	Not NN nor NNS
JJ	JJ	Not NN or NNS
NN or NNS	JJ	Nor NN nor NNS
RB, RBR, or RBS	VB, VBD, VBN, VBG	anything

How to measure polarity of a phrase?

- Positive phrases co-occur more with “excellent”
- Negative phrases co-occur more with “poor”
- But how to measure co-occurrence?

Pointwise Mutual Information

- PMI between two words:
 - How much more do two words co-occur than if they were independent?

$$PMI(word_1, word_2) = \log_2 \frac{P(word_1, word_2)}{P(word_1)P(word_2)}$$

How to estimate PMI

- $P(\text{word})$ estimated by $\text{hits}(\text{word})/N$
- $P(\text{word}_1, \text{word}_2)$ by $\text{hits}(\text{word}_1 \text{ NEAR } \text{word}_2)/N^2$

$$PMI(\text{word}_1, \text{word}_2) = \log_2 \frac{\text{hits}(\text{word}_1 \text{ NEAR } \text{word}_2)}{\text{hits}(\text{word}_1)\text{hits}(\text{word}_2)}$$

Does phrase appear more with “poor” or “excellent”?

Phrase	POS.tags	Polarity
online service	JJ NN	2.8
online experience	JJ NN	2.3
direct deposit	JJ NN	1.3
local branch	JJ NN	0.42
...		
low fees	JJ NNS	0.33
true service	JJ NN	-0.73
other bank	JJ NN	-0.85
inconveniently located	JJ NN	-1.5
<i>Average</i>		<i>0.32</i>

Phrase	POS.tags	Polarity
direct deposits	JJ NNS	5.8
online web	JJ NN	1.9
very handy	RB JJ	1.4
...		
virtual monopoly	JJ NN	-2
lesser evil	RBR JJ	-2.3
other problems	JJ NNS	-2.8
low funds	JJ NNS	-6.8
unethical practices	JJ NNS	-8.5
<i>Average</i>		<i>-1.2</i>

Phrases from a thumbs-up (positive) review

Phrases from a thumbs-down (negative) review

Results of Turney algorithm

- 410 reviews from Epinions
 - 170 (41%) negative
 - 240 (59%) positive
- Majority class baseline: 59%
- Turney algorithm: 74%
- Phrases rather than words
- Learns domain-specific information

Using WordNet to learn polarity

- WordNet: online thesaurus
- Create positive (“good”) and negative seed-words (“terrible”)
- Find Synonyms and Antonyms
 - Positive Set: Add synonyms of positive words (“well”) and antonyms of negative words
 - Negative Set: Add synonyms of negative words (“awful”) and antonyms of positive words (“evil”)

Learning lexicons in summary

- Advantages:
 - Can be domain-specific
 - Can be more robust (more words)
- Intuition
 - Start with a seed set of words ('good', 'poor')
 - Find other words that have similar polarity:
 - * Using “and” and “but”
 - * Using words that occur nearby in the same document
 - * Using WordNet synonyms and antonyms

Supervised Methods

Document sentiment classification

- Classify a document (e.g., a review) based on the overall sentiment of the opinion holder
 - Classes: Positive, negative (possibly neutral)
- An example review:
 - “I bought an iPhone a few days ago. It is such a nice phone, although a little large. The touch screen is cool. The voice quality is great too. I simply love it!”
 - Classification: positive or negative?
- It is basically a text classification problem

Sentence sentiment analysis

- Classify the sentiment expressed in a sentence
 - Classes: positive, negative, neutral
 - Neutral means no sentiment expressed
 - * “I believe he went home yesterday.”
 - * “I bought a iPhone yesterday”
- But bear in mind
 - Explicit opinion: “I like this car.”
 - Fact-implied opinion: “I bought this car yesterday and it broke today.”
 - Mixed opinion: “Apple is doing well in this poor economy”

Features for supervised learning

- The problem has been studied by numerous researchers.
- Key: feature engineering. A large set of features have been tried by researchers. E.g.,
 - Terms frequency and different IR weighting schemes
 - Part of speech (POS) tags
 - Opinion words and phrases
 - Negations
 - Syntactic dependency

Sentiment classification in movie reviews

- Polarity detection:
 - Is an IMDB movie review positive or negative?
- Data: Polarity Data 2.0:
 - <http://www.cs.cornell.edu/people/pabo/movie-review-data>

Basic steps

- Pre-processing and tokenization
- Feature representation (DTM)
- Feature selection
- Classification

Sentiment tokenization issues

- Deal with HTML and XML markup
- Twitter mark-up (names, hash tags)
- Capitalization (preserve forwords in all caps)
- Phone numbers, dates
- Emoticons
- Useful code:
 - Christopher Potts sentiment tokenizer
 - Brendan O'Connor twitter tokenizer

Potts emoticons

[<>]?	# optional hat/brow
[:=8]	# eyes
[\-o*\']?	# optional nose
[\)\]\(\[dDpP/\:}\{\@\\ \\]	# mouth
	#### reverse orientation
[\)\]\(\[dDpP/\:}\{\@\\ \\]	# mouth
[\-o*\']?	# optional nose
[:=8]	# eyes
[<>]?	# optional hat/brow

Extracting features for sentiment classification

- How to handle negation
 - I didn't like this movie vs
 - I really like this movie
- Which words to use?
 - Only adjectives
 - All words
 - * All words turns out to work better, at least on this data

Negation

Add NOT_ to every word between negation and following punctuation:

didn't like this movie , but I

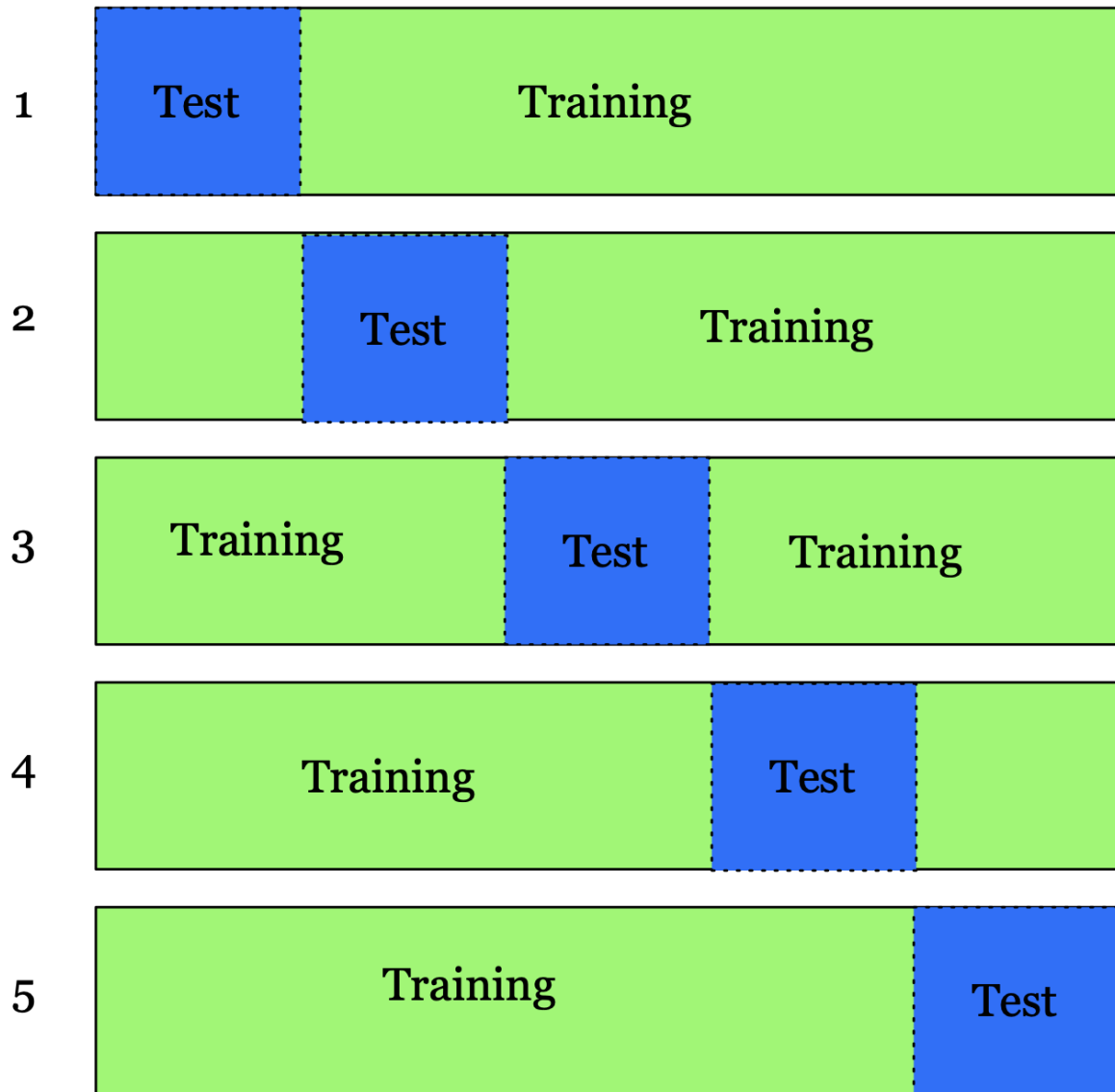


didn't NOT_like NOT_this NOT_movie but I

Cross-Validation

- Break up data into 10 folds
 - (Equal positive and negative inside each fold?)
- For each fold
 - Choose the fold as a temporary test set
 - Train on 9 folds, compute performance on the test fold
- Report average performance of the 10 runs

Iteration



Supervised sentiment analysis

- Using all words works well for some tasks
- Finding subsets of words may help in other tasks
 - Hand-built polarity lexicons
 - Use seeds and semi-supervised learning to induce lexicons
- Negation is important

Other Challenges in SA

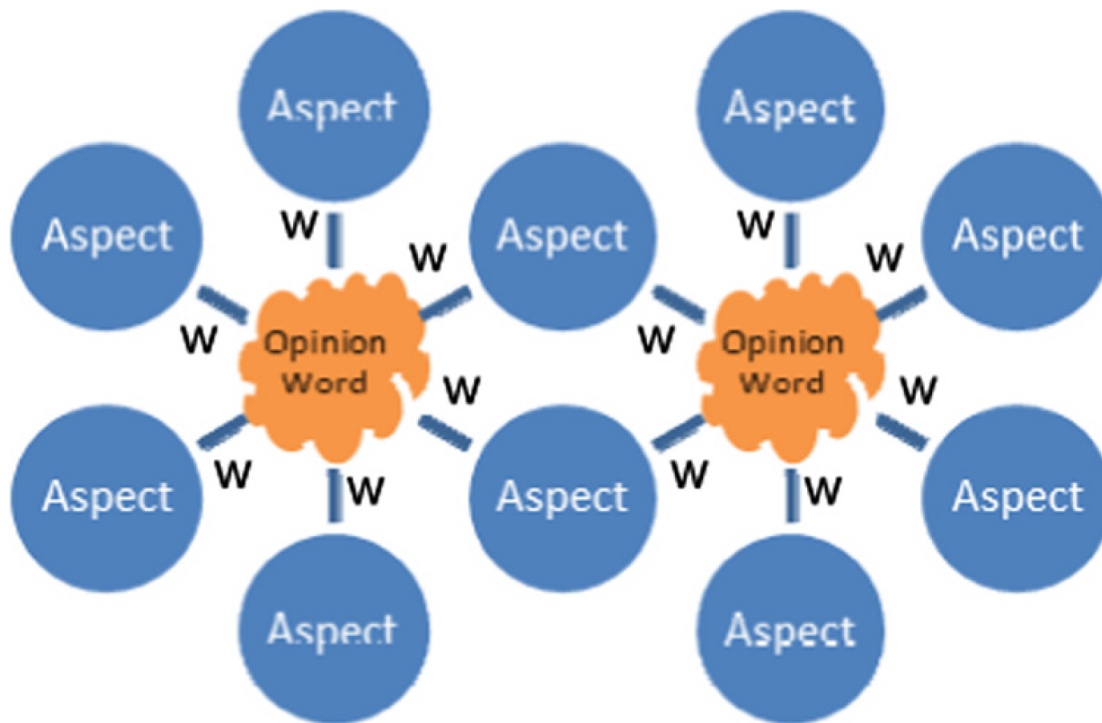
Explicit and implicit aspects | (Hu and Liu, 2004)

- Explicit aspects: Aspects explicitly mentioned as nouns or noun phrases in a sentence
 - “The picture quality is of this phone is great.”
- Implicit aspects: Aspects not explicitly mentioned in a sentence but are implied
 - “This car is so expensive.”
 - “This phone will not easily fit in a pocket.”
 - “Included 16MB is stingy.”

Implicit aspects | Bagheri et al. 2013

An implicit aspect should satisfy the following conditions:

- The related aspect word does not occur in the review sentence explicitly.
- The aspect can be discovered by its surrounding words (e.g. opinion words) in the review sentence.



co – occurrence(aspect, opinion word)

$$= \log \left(\frac{w_{aspect, opinion \ word}}{degree_{aspect} * degree_{opinion \ word}} + \varepsilon \right)$$

Some interesting sentences

- Trying out Chrome because Firefox keeps crashing.
 - Firefox - negative; no opinion about chrome.
 - We need to segment the sentence into clauses to decide that “crashing” only applies to Firefox(?).
- But how about these
 - I changed to Audi because BMW is so expensive.
 - I did not buy BMW because of the high price.
 - I am so happy that my iPhone is nothing like my old ugly phone.

Some interesting sentences (contd)

- Conditional sentences are hard to deal with (Narayanan et al. 2009)
 - If I can find a good camera, I will buy it.
 - But conditional sentences can have opinions
 - * If you are looking for a good phone, buy Nokia
- Questions are also hard to handle
 - Are there any great perks for employees?
 - Any idea how to fix this lousy Sony camera?

Some interesting sentences (contd)

- Sarcastic sentences
 - What a great car, it stopped working in the second day.
- Sarcastic sentences are common in political blogs, comments and discussions.
 - They make political opinions difficult to handle

Summary

Summary

- Sentiment analysis
- Lexicon-based methods
- Learning-based methods

Practical 9