

Computational approaches to emotions in language

Diana Santos

ILOS

d.s.m.santos@ilos.uio.no

June 2015, day 3



Opinion mining and sentiment analysis

From Pang & Lee (2008)

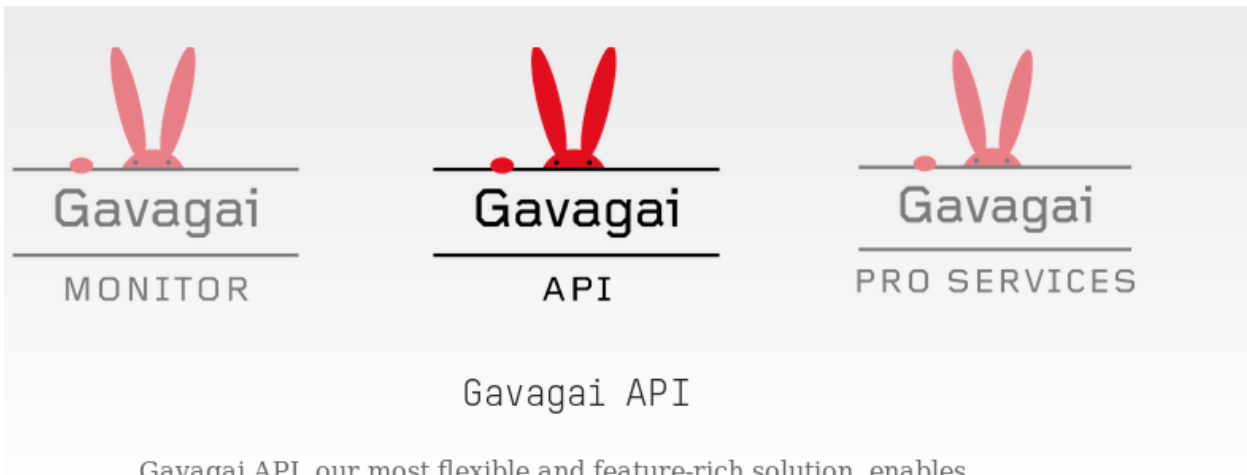
- Opinion search: products, political arguments, individuals' reputations...
- Difficulties:
 - query classification: is the document about the product, the person?
 - what part of the document conveys the opinion vs. the facts?
 - overall opinion?
 - How to present the results?

Different applications and uses

- web search targeted towards reviews
- help (or data) to recommendation systems
- detection of “flames” in on-line communities
- sensitive content for ads placement
- help information retrieval by discarding (too) subjective statements
- opinion-oriented questions
- summarization of multiple viewpoints
- more intelligent citation analysis
- analysis of literary reputation

Different applications and uses (cont'd)

- improvement of “human-computer interaction”
- business and government intelligence (finding what competitors are saying or how they are being reviewed)
- metastudies in sociology: how do opinions diffuse?



Gavagai API, our most flexible and feature-rich solution, enables you to add online monitoring as well as next generation text analytics to your products and internal applications.

Use it to keep track your brand, products or sectors in news and social media, import your own text data for Multi-Dimensional Sentiment analysis, and generate summaries of vast sets of texts written in multiple languages using our proprietary Multi-Document Summarization solution.

If you need something a bit different, **contact us** and we'll be happy to customise a plan for you.

Kinds of tasks

- Sentiment polarity detection
- Classification (sentiment classification task)
- Extraction
- (Viewpoint) summarization

Sentiment polarity

Sentiment polarity: attribute positive or negative polarity to a document as a whole.

- Opinions can be subtle – so even human beings may be in doubt, if you read ironical or full of allusions (comparisons) posts, or even if you fail to understand the larger context
- The lexicon can be misleading
- One can be citing or even denying other opinions, so opinion attribution is extremely relevant
- Emoticons make the interpretation even more complicated ;-)
- As in any rating system, you have to know the style of the rater: for him, this was a huge praise, for her, everyone was “wonderful”
- One typical text strategy is finishing by a opposite (but definitive) conclusion.

This film should be brilliant.... However, it can't hold up.

- Kind of text/application: patient-doctor interviews in hospital are different from comments on new software or arguments on Wikipedia

Classification (sentiment classification task)

- Is a text reflecting an opinion, or just facts?
- Does text consist of good news, or bad news? (Or better, is it presented by the writer as good news or bad news?)
- Are there properties that are obviously good (or bad)? Such as *long battery life*, *large rooms*, etc. (Of course *long* and *large* are not necessarily good.)
- Some words are absolutely evaluative, though, such as: *only*, *too*, *even*, or openers like *I'm afraid*:
 - the battery only lasts two hours
 - the movie is too long
 - even Einstein had his failures

In reviews, one often has pros and cons, and this should be extracted separately, for multidocument summarization, for example.

Also, there are reviews which are more objective than others:

- *I love it!*
- *This camera has the following advantages over X and Y: a, b, c*

What about rating in an ordinal scale? Are the comments commensurate? Should one separate between “neutral” and no opinion (“lack of opinion”)? One other issue: discard off-topic passages. Is topic always the same in one document/text?

Finally: whose opinion it is? Especially relevant in political applications.

Evaluation to foster progress

The 2006 Blog track at TREC: Extracting subjectivity from blogs, or opinion retrieval task (Ounis et al., 2006).

- 100,000 blog feeds, 11 weeks, + collections of top blogs + spam, yielding 3.2 million documents (blog + comments are one unit)
- topics were created by having assessors interpretate real logs to search engines
- spam blogs detection
- distinguish between: no-opinion, positively or negatively opiated documents, or both

TASS 2014: Aspect based sentiment analysis: reputation (“tracking, investigating and reporting an entity’s actions and other entities’ opinions about those actions”, in Spanish) (Villena Román et al., 2015)

- 68,000 tweets from 2011-2012, semi-automatically tagged with polarity and topic
- tweets (Social TV corpus), 2773 tweets manually tagged for aspects related to a particular match: fans, referee, authorities, coach, team, players, etc. and with polarity.

Agreement detection

Do two texts (or two authors) agree on a particular issue?

This is specially relevant for systems that study political opinions, but can also be conceived as a subpart of an extraction system.

Examples: Israeli vs. Palestinian viewpoints; liberal, conservative, libertarian; political parties's cahracterizations, etc.

Task: to automatically answer questions on "what is X's perspective on Y"?

Mood detection

Spoken dialogue systems: to detect whether the user is annoyed is relevant. And to detect humor and produce it.

To detect whether a document is "high-brow" vs. popular...

And to detect deceptive language, for example in e-mail (Zhou et al., 2008)

To detect mood in blogs...

Interesting findings

- Hapax legomena are indicators of subjectivity
- Presence more important than frequency
- (Gradable) adjectives are relevant, but also (emotion) verbs or nouns (*hope, gem*)
- Taking negation into consideration is hard

History and data

(according to Pang & Lee (2008))

- In the 1990s, there started to appear “organized collections of opinionated documents”, such as financial-news discussion board, or Epinion, and from then on... ratings, Web 2.0, etc.
- and large lexicons such as WordNet
- active lines:
 - text categorization (generally by machine learning algorithms)
 - lexicon induction

Food: Delicious, Art: beautiful, thought provoking, Fashion: elegant, well-cut.

- “Go read the book” - is good for book reviews, bad for movie reviews; “unpredictable” is good for plots but bad for cars; experienced...
- size of reviews/opinions implies different classifiers altogether.
- temporally dependence is important: different time periods show different lexicons (Read, 2006)
- cross-lingual adaptation: *deilig* in Norwegian can be used for art and fashion, but *flink* is notoriously difficult to translate into Portuguese, for example. Different cultures praise and criticize in different ways

... and topic dependence

A different issue is the topic

- Two pass approaches, or simultaneous modelling?
- On-topic and off-topic... what about multitopic documents?
- Some times the topics one is interested in are determined, other times it is relevant to see what topics people are emotional about.

A typical unsupervised approach

- 1 First, create an unsupervised sentiment lexicon from a corpus of opinionated texts, assigning “semantic orientation” to words.
 - studying oppositions (but or and?)
 - mutual information, conjunction
- 2 Using it to determine how positive (or negative) each text is
 - distinguish between prior polarity and contextual polarity
- 3 Possibly bootstrap to get a better lexicon...

More complex approaches: using relationships

Not “bag of words”, look at relationships

- among sentences inside a document: All sentences are labelled as subjective or objective through graph methods (finding an s/l cut in the graph).
- among participants: agreement towards other posts or participants in a debate (when you answer, are you antagonistic or reinforcing?)
When you quote, do you quote what you like or what you dislike? (Malouf & Mullen, 2008)
- among product features: collectively adjust the sentiment labels
- among classes: if one has a scale, not just positive, negative and neutral... differences between ratings are relevant, intensity classification is also interesting

More complex approaches: discourse structure

- include location information in the computation. The most important places are
 - end for real texts,
 - beginning for summaries
- compute the trajectory of local sentiments in a text

More complex approaches: Faceted extraction

Sometimes one is interested in opinions/evaluation of different aspects of a particular product (or person):

- What kinds of features do people have opinions about?
- Is there a difference between “properties” and “parts”? For a scanner, size is a property, cover is a part (Popescu & Etzioni, 2005)
- Which opinions, and which weigh more? It depends on the user/reader?
- Can (the single mention of) objective properties represent an opinion/evaluation?

The identification of the holder of the opinion... may be the most important thing about an opinion.

- it is usually not easy to trace down the exact sources
- co-reference resolution on the same individual can be tricky (Stoyanov & Cardie 2006)
- reporting what the others say (or mean) includes a great deal of subjectivity
- reporting other peoples' feelings can even be worse

Summarization of emotion/opinion content

And now to aggregation of all we extracted so far.

- single document summarization
 - single passage...
 - incorporation of sentiment flow
 - all subjective sentences
 - visual summaries from information extraction
 - graph summaries
 - how to link evidence to opinion? Vexing question (Wiebe et al., 2004)
- multi-document summarization
 - agreement on a rating may be due to different reasons/grounds
 - contrary to discarding redundancy, the most important is to measure it, as well as the degree of conflicting opinions.
 - summaries as evaluative arguments created by a natural language generator
 - the headlines of the most positive and the most negative opinions
 - or, representative sample of opinions
 - summary statistics: thermometer, colour, size, for bounded cases
 - summary statistics: how to distinguish between a normal distribution or a two mode one if you just give the average?

Opinion observer...

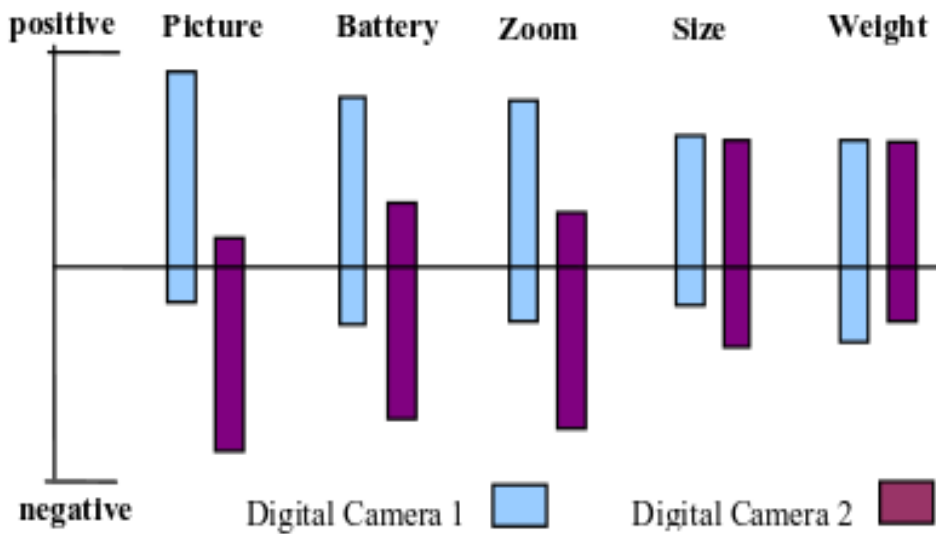


Figure 1: Visual comparison of consumer opinions on two products.

From Liu et al. (2005).

From Morinaga et al. (2002), next slide

Degree of association with common phrases

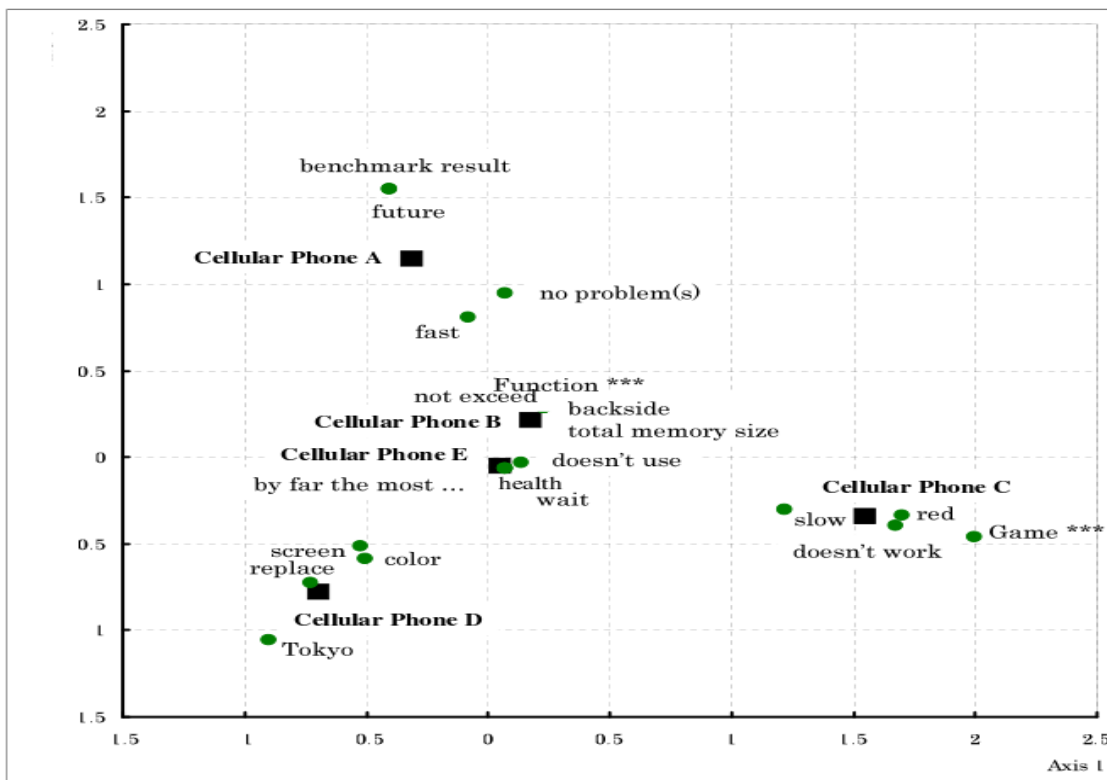


Figure 5: Positioning Map for Five Cellular phones and Their Extracted Characteristic Words

- Order of presentation of the results
- Tracking changes in opinions
- Reputation/awareness analysis (“all publicity is good publicity”)

Quality of the reviewers – and of the reviews

- What is “high status” vs. “low status” literature?
- Rating the reviews: X of Y found this useful...
- Utility is not directly related to quality
- Automatic quality ranking
- Opinion-spam detection
- Detection of influential or trustworthy authors (Pinski & Narin 1976, Song et al. 2007)

- history of market manipulation through disinformation
- reputation manipulation
- are reviews relevant? the economic impact of reviews.

Other systems and approaches

Since Pang & Lee's overview was done in 2008, some pointers to modern challenges and solutions:

- studies of emotions in new media (tweets etc)
- studies of emotions/reputation together with other modes (dialogue systems, gestures, faces)
- studies of generation of emotions (in language or in avatars)
- applications like intelligent houses, medicine helpers, pedophilia detection, etc.