Sentiment analysis in the lexicon and in the discourse

Josef Ruppenhofer

July 1, 2009
Overview

- What do you mean, Sentiment?
- Sentiment denotation v. invited inference
- Attribution
- Variety
- Context dependence
- (Non-)Compositionality
- Automatic processing of sentiment expressions
The core cases

- Expressions of judgment/evaluation/emotion
  - Wow, this is my 4th Olympus camera.
  - Staley declared it to be one hell of a collection.
Why do sentiment analysis?

- Fact-based question answering
  - Q: When is the first day of spring in 2007?
  - Q: Does the US have a tax treaty with Cuba?

- Opinion question answering
  - Q: What is the international reaction to the re-election of Robert Mugabe as President of Zimbabwe?

- Product Review Mining
  - Is that restaurant/book/movie/laptop any good?
A word on terminology

- There are quite a few terms in use in the area of sentiment analysis
  - Subjectivity analysis
  - Opinion analysis
  - Sentiment analysis
- Often they are used interchangeably but in some cases they are meant to evoke a particular theoretical background or history
- Generally, subjectivity analysis takes the broadest view of the object of study
Less central cases

- **Belief**
  - Most voters believe that he’s not going to raise their taxes

- **Suasion/Arguing**
  - One more word out of you and I’ll have you investigated.
  - We can enhance our sales, if we add speech recognition.
Inspirations of Subjectivity analysis

- literary theory
  - Subjective sentences (*Unspeakable Sentences*, Ann Banfield, 1982)
    1. "What are you doing in here?"
    2. *Suddenly she [Zoe] was furious with him [Joe].*
    3. "Spying, of course."
    4. "Well of all dumb things!"
    5. I thought you ran away.
    6. *Joe Bunch was awful.*
  - Subjective sentences take the point of view of particular character.
  - Subjective sentences have subtle but identifiable grammatical properties.
    - A few drops of rain were **now** falling.
    - He said he would go home **now/then**.
Inspirations of Subjectivity analysis

- linguistic theory
  - Private states (Quirk et al. 1985)
  - The private subtype consists of verbs expressing intellectual states such as belief and intellectual acts such as discovery. These states and acts are ‘private’ in the sense that they are not observable: a person may be observed to assert that God exists, but not to believe that God exists.
Inspirations of Subjectivity analysis

Some kinds of linguistic choices that depend on the situated experience of a particular person are ignored by subjectivity analysis.

- **Spatial, temporal, and person deixis**
  - Turn left v. Turn west
  - yesterday v. October 5

- **Fictive construals of motion**
  - Trees flew past us
  - The road winds through the mountains.
Information Targeted By Sentiment Analysis

- **Source** (whose psychological/emotional state; aka Opinion Holder)
- **Target** (who or what it’s about; aka Topic)
- **Polarity [aka Valence]** (positive or negative; pro or contra; ...)
- **Attitude** (Sentiment; Arguing; ...)
- **Intensity** (loathe > hate > dislike > disprefer > not like very much)
Sentiment viewed linguistically

- Some interesting features of sentiment expression
  - denotation v. invited inference
  - attribution problem
  - variety
  - frequent dependence on context
    - subjective or not
    - polarity
  - non-compositionality
Opinions v. polar facts

- Some statements invite emotional appraisal but do not explicitly denote appraisal.
- While such polar facts may in a particular context seem to have an obvious value, their evaluation may be very different in another one.
XP not Vista

- **Pros:** Price/Value. XP OS NOT VISTA! Screen good even in bright daylight. Easy to access USB, lightweight.
- **Cons:** machine has the older XP not vista.
Attribution
Attribution

- I like ice-cream.
- Bill says that Sue hates Harry.
- In Yugoslavia’s dying days, Slobodan Milosevic slandered Mesic as Ustashe and blocked his presidency.
- She told her she is going to marry that idiot.
Attribution: types of sentiment expressions

- References to private states
  - His *alarm* grew
  - He was boiling with *anger*
- References to speech or writing events expressing private states:
  - UCC leaders roundly *condemned* the Iranian President
  - "He’s a quack", He *said*.
- Expressive subjective elements
  - He would be quite a *catch*
  - He’s a *quack*. 
“The US fears a spill-over”, said Xirao-Nima, a professor of foreign affairs at the Central University for Nationalities.
Attribution and the level of analysis

Research commonly distinguishes between document, sentence/clause, and expression-level sentiment analysis. In particular tasks, attribution is (often) ignored.

- product review mining, e.g. on Amazon
- movie reviews
- newspaper editorials
- blogs
Variety
The list of linguistic expressions conveying sentiment is extremely long

- darling
- gem
- sweetheart
- marvel
- dimwit
- blockhead
- idiot
- ...
Variety

- Sentiment can be expressed by many different linguistic means
  - your husband is a darling
  - your dear husband
  - your husband, that darling,
  - I like your husband
  - your darling of a husband
“I'm not spinning—I'm contextualizing.”
Context dependence: Presence of subjectivity

▶ Word senses
  ▶ Neutral: Perhaps, ironically, it has something to do with the idea that, historically, the **liberal** arts are connected to the cultivation of gentlemen and gentlewomen.

▶ Paradigmatic choices
  ▶ Ask Daddy if he can take you to school.
  ▶ Ask **that man** if he can take you to school. (where “that man” is the addressee’s father and the speaker’s husband)
  ▶ Könnten Sie mir helfen?
  ▶ Könntest **Du** mir helfen? (rude if addressed to an unfamiliar adult in a formal context)
Context dependence: constructional influences

- NP be hell: That plane ride was hell.
- a hell of a NP: Are You Ready for a Hell of a Good Time?
- one hell of a NP: I am having one hell of a time trying to consolidate and organize my music collection.
Context dependence: Presence of subjectivity

- Associations are not stable across time
  - for word senses (E knave - G Knabe)
  - for paradigmatic choices
    - estate tax v. death tax
    - French fries v. freedom fries
Context dependence: Presence of subjectivity

- Associations differ between speakers (ideology)
  - liberal
    - The liberal elite have clearly fallen out of touch with the American voter.
    - I'm a Fighting Liberal, and Proud of What that Means
Context dependence: Polarity

- Some constructions mark presence of subjectivity but leave polarity open
  - the next N.N.
    A leading black conservative says Democratic presidential frontrunner Barack Obama is **not the next John F. Kennedy**. Instead, says Ken Blackwell, he’s **the next George McGovern** and ”it’s time people learned the facts.”!
  - Person be such a N
    - **You’re such a good kisser**
    - You, **you’re such a big star** to me.
    - Oh, Lisa, **you’re such a girl!**
Idiomaticity

“Good news. The test results show it’s a metaphor.”
Idiomaticity/non-compositionality

- a lot of subjective language consists of MWEs
- but none of the components may be sentiment evoking in isolation
  - be not a altar boy ("Your son is no altar boy," the mother is told.)
  - have had it with N (We’ve had it with the attacks on our patriotism)
  - be the third rail of (Antidumping is the "third rail" of U.S. trade politics)
  - Go Down Like A Lead Balloon
Idiomaticity/non-compositionality

- and even if there are such components, they may not add up or be hard to integrate
  - goody two-shoes (Everyone despises a Miss Goody Two Shoes, and Isidingo’s Thandi has to be the epitom of goody two shoes.)
  - love to death (I just love her to death.)
  - no love lost (There is no love lost between them)
Idiomaticity/non-compositionality

- Subjective expressions can be embedded under, or modified by, other expressions with the effect that their polarity and/or intensity is modulated.
  - *I have faith* in the president.
  - *I lack faith* in the president.
  - *I have hardly any faith* in the president.

- In sentiment analysis, people often describe the effect of words like *lack* on *faith* as **valence shifting** (Polanyi & Zaenen 2005).

- It would be nice to compose the sentiment values of linguistic expressions (Moilanen & Pulman 2007) ...
Idiomaticity/ non-compositionality

... but despite many simple cases that is a difficult challenge

► Cheers to Timothy Whitfield for the wonderfully horrid visuals.
► I don’t like reggae. I love it
► He’s never met a dictator he didn’t like.
► Has he ever seen a tax he didn’t like?
► I doubt she’s ever come across a military spending bill she didn’t like.
► Name me even just one tax increase she didn’t like!
► At least, she’s never held up a tax bill she didn’t like.
Sentimental inferences

- Often we can infer sentiments beyond those that are directly stated
  - I am **glad** that Chelsea lost.
    - Negative towards Chelsea: I hate that team.
    - Neutral: I want Arsenal to become champions.
    - Positive: This way they don’t have so many games and can concentrate on the championship.

- Interestingly, the simple inference (Negative towards Chelsea) is correct in the majority of instances.
Sentiment is in the grammar

- Some comparisons of equality **must** be interpreted metaphorically/evaluatively
  - Tom is as fat as Bill.
  - Tom is as fat as a cow.
  - *Tom is __ fat as Bill*
  - Tom is __ fat as a cow.

- There are idiomatic constructions dedicated to expressing subjectivity:
  - Since you’re going to be there a while, why not go and chit-chat? (why+VP)
  - Customer to Waiter: ”What is that fly doing in my soup?” Waiter. ”Looks like the back stroke. (WXDY)
Using context to process subjective expressions

- for the purpose of lexical analysis
  - use distributional hypothesis to find expressions that
    - are also subjective/objective (Wiebe 2000)
    - given subjective status, have same (prior) polarity as seed words
  - use contextual coherence to learn polarity during lexicon acquisition (Kanayama et al. 2006)
  - use syntagmatic co-occurrence with seed terms as measure of polarity (Turney 2002, Hatzivassiloglou & McKeown 1997)

- for analysis of particular uses
  - use contextual coherence to decide polarity of expressions with flexible polarity
  - look for valence shifters that may influence polarity of expressions on particular uses (Wilson 2008)
Non-contextual methods

- Analyzing dictionaries, esp. Wordnet (Gyamfi, Wiebe, & Mihalcea, R. 2009)
- Acquire subjective expressions (+polarity) via aligned-corpora (Banea, Mihalcea, & Wiebe, J. 2008)
Resources

- Jan Wiebe’s website
  - bibliography
  - Opinionfinder, including dictionaries
  - MPQA corpus
- Esuli and Sebastiani’s SentiWordNet
- Bing Liu, who has a product mining angle