Automatische Sentimentanalyse zwischen Hotel und Parlament

Manfred Stede
Universität Potsdam

Digital * Humanities im Gespräch / FU Berlin / 6. Dezember 2018

Overview

• Sentiment analysis: Introduction, Terminology
• One system: SO-CAL
• Application: Political text
• Application: Narrative
• Viewpoint: Sentiment / DH
One Bolzano hotel @ tripadvisor.com

“Mixed feelings”
★★★★ Reviewed 3 days ago  via mobile

Firstly the good points .... We had a very large room with fantastic bathroom and walk in closet. There was a good breakfast selection and possible to eat outside. There was a pretty garden with an outside bar and it was nice to sit outside after dinner. Location was excellent and a couple of minutes walk to the main square ...
More

Helpful? Thank AJG7

Near-Synonyms?

• Opinion mining
• Sentiment analysis
• Subjectivity analysis
Subjectivity

- I don’t like this wine.
- There is a cat on the mat.
- I’m dizzy.
- Peter adores Barack Obama.
- I don’t think that Trump can win the election.
- Last night I met this really nice musician.
- Hooray!
- That’s probably a dromedar, not a camel.

Subjectivity

- The linguistic expression of somebody’s opinions, sentiments, emotions, evaluations, beliefs, speculations (Wilson/Wiebe: MPQA guidelines)

- Private state: state of a speaker/writer that is not open to objective observation or verification


- Automatic subjectivity analysis classifies content as objective or subjective
Subjectivity

- **Sentiment**: an attitude or feeling (not necessarily directed toward something)
- **Opinion**: an evaluation of something (necessarily directed)

=> Sentiment analysis and Opinion mining overlap, but there can be sentiment analysis that does not mine opinions (e.g., capture the general mood in the newspapers)
- **In practice**, most automated systems reduce evaluation to **polarity**

Text-level sentiment analysis

- Firstly the good points .... We had a very large room with fantastic bathroom and walk in closet. There was a good breakfast selection and possible to eat outside. There was a pretty garden with an outside bar and it was nice to sit outside after dinner. Location was excellent and a couple of minutes walk to the main square.

=> **positive**
Text-level sentiment analysis

• Firstly the good points .... We had a very large room with fantastic bathroom and walk in closet. There was a good breakfast selection and possible to eat outside. There was a pretty garden with an outside bar and it was nice to sit outside after dinner. Location was excellent and a couple of minutes walk to the main square.

• => positive

The most basic approach

Lexicon of words with prior polarity

– excellent, fantastic, good, nice, ...
– boring, terrible, uncool, ugly, ...

Simple Method:

• Preprocess the text (stemming, lemmatization)
• Count positive and negative words in document
Units for sentiment analysis

- **Text**
  - assume it has one topic and one overall orientation
- **Paragraph**
  - likewise; can compute text orientation afterward
- **Sentence**
  - likewise; can compute para orientation afterward
- **Phrase**
  - can capture things like *While the breakfast was good, I couldn’t stand dinner*

Extensions (1): Polar facts

- Firstly the good points .... *We had a very large room with fantastic bathroom and walk in closet. There was a good breakfast selection and possible to eat outside. There was a pretty garden with an outside bar and it was nice to sit outside after dinner. Location was excellent and a couple of minutes walk to the main square.*
Extensions (2): Aspects

• Firstly the good points .... We had a very large room with fantastic bathroom and walk in closet. There was a good breakfast selection and possible to eat outside. There was a pretty garden with an outside bar and it was nice to sit outside after dinner. Location was excellent and a couple of minutes walk to the main square.
Extensions (3): Fine-grained analysis

*While the breakfast was good, I couldn’t stand dinner*

- Words with prior polarity
- Intensifiers/diminishers
- Source
- Target
Extensions (4): Entity-level sentiment

Roger Federer won the match against Nadal, who had been fervently supported by the audience.

What is good or bad for whom?

Roger Federer won the match against Nadal, who had been fervently supported by the audience.
Extensions (4): Entity-level sentiment

Roger Federer won the match against Nadal, who had been fervently supported by the audience.

What is good or bad for whom?

Roger Federer won the match against Nadal, who had been fervently supported by the audience.
I can’t say that I enjoyed my stay at the Belvedere Hotel. Other reviewers said it’s a great place, but my impression was otherwise. Neither was the food particularly good, nor did we consider the location very convenient. Just a standard place to live for a day, that’s it.
Extensions (5): Contextual polarity

I can’t say that I enjoyed my stay at the Belvedere Hotel. Other reviewers said it’s a great place, but my impression was otherwise. Neither was the food particularly good, nor did we consider the location very convenient. Just a standard place to live for a day, that’s it.

Overview

• Sentiment analysis: Introduction, Terminology
• **One system**: SO-CAL
• Application: Political text
• Application: Narrative
• Viewpoint: Sentiment / DH
SO-CAL

• Semantic Orientation CALculator
• Text-level polarity analysis (graded)

• Selling points:
  – Use crowdsourcing in building a lexicon
  – Rule-based approach to contextual polarity (with some new ideas)
  – Achieves good level of domain-neutrality


Size
2252 adjectives
1142 nouns
903 verbs
745 adverbs

Words collected from 500 movie and product reviews (8 categories, balanced for pos and neg)

Manually ranked on -5 .. 5 scale: prior polarity and strength

<table>
<thead>
<tr>
<th>Word</th>
<th>SO Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>monstrosity</td>
<td>-5</td>
</tr>
<tr>
<td>hate (noun and verb)</td>
<td>-4</td>
</tr>
<tr>
<td>disgust</td>
<td>-3</td>
</tr>
<tr>
<td>sham</td>
<td>-3</td>
</tr>
<tr>
<td>fabricate</td>
<td>-2</td>
</tr>
<tr>
<td>delay (noun and verb)</td>
<td>-1</td>
</tr>
<tr>
<td>determination</td>
<td>1</td>
</tr>
<tr>
<td>inspire</td>
<td>2</td>
</tr>
<tr>
<td>inspiration</td>
<td>2</td>
</tr>
<tr>
<td>endear</td>
<td>3</td>
</tr>
<tr>
<td>relish (verb)</td>
<td>4</td>
</tr>
<tr>
<td>masterpiece</td>
<td>5</td>
</tr>
</tbody>
</table>

Manfred Stede (Uni Potsdam)
Lexical ambiguity

• **Sense** ambiguity: sometimes resolved via PoS
  – *plot*: neutral noun, negative verb
  – *novel*: neutral noun, positive adjective

• **Connotation** ambiguity: „resolved“ by averaging
  – The teacher *inspired* her students to pursue their dreams.
  – This movie was *inspired* by true events.

Derivations

• Some nouns **derived automatically** from verb dictionary, but strength can change
  – *exaggerate*: -1
  – *exaggeration*: -2
  – also: *complicate / complication*, etc

  – (hypothesis: general trend?)
Derivations (2)

• Adverb dictionary built from adjectives, by –ly matching
• sometimes value needs to be corrected: *essential / essentially*

<table>
<thead>
<tr>
<th>Word</th>
<th>SO Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>excruciatingly</td>
<td>-5</td>
</tr>
<tr>
<td>inexcusably</td>
<td>-3</td>
</tr>
<tr>
<td>foolishly</td>
<td>-2</td>
</tr>
<tr>
<td>satisfactorily</td>
<td>1</td>
</tr>
<tr>
<td>purposefully</td>
<td>2</td>
</tr>
<tr>
<td>hilariously</td>
<td>4</td>
</tr>
</tbody>
</table>

Context: Intensification

• *amplifiers* *(very) downtoners (slightly)*
• Polanyi/Zaenen 06, Kennedy/Inkpen 06: add and subtract values
• BUT: degree of intensification should depend more on the word intensified => multiply
• 177 intensifiers in the lexicon

3 x (100 + 25%) x (100 + 15%) = 4.3

really very good₃
Context: Negation

- The acting was not very good.

- Some negators appear at long distance
  - Nobody gives a good performance in this movie.

- Strategy: Look backwards until a clause boundary (punctuation or connective) is reached
  - I don’t think this will be a problem.

Context (2): Negation - value change

- One approach: polarity flipping (e.g., Choi/Cardie 08)
- Problems
  - excellent: +5
  - not excellent: -5 ??
  - atrocious: -5

- => Use polarity shift (+/-4) rather than flip
  - The food is not terrific (5 – 4 = 1) but not terrible (-5 + 4 = -1) either.
  - It’s not a spectacular (5 - 4 = 1) film.
Context (3): Irrealis blocking

- For kids, this movie could be one of the best of the holiday season.
- I thought this movie would be as good as the Grinch, but unfortunately it wasn’t.
- Implementation: ignore polar words in the scope of an irrealis marker (scope: heuristic)
  - modals
  - conditionality
  - NPIs (any, anything, ..)
  - questions
  - material in quotes
  - certain verbs (doubt, expect, ..)
- This should have been a great movie. (3 -> 0)

Example output

- The food was wonderful. => 5.0
- The food was particularly good. => 3.9
- The food was good. => 3.0
- The food was not bad. => 2.0
- The food was OK. => 2.0
- The food was not particularly good. => -0.1
- I didn’t really enjoy the food. => -0.4
- I didn’t enjoy the food. => -1.0
- The food was bad. => -3.0
Evaluation: Lexicon complexity

- Use not/recommended value of the review: >0 / <0
- 3 variants of the approach
  - Simple: only 2/-2 values and 1/-1 intensification (Polanyi/Zaenen 06)
  - Only-Adj: use only adjectives
  - One-Word: don’t use multi-word expressions
  - Full: complete system as described

<table>
<thead>
<tr>
<th>Dictionary</th>
<th>Epinions 1</th>
<th>Epinions 2</th>
<th>Movie</th>
<th>Camera</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>76.75</td>
<td>76.50</td>
<td>69.79*</td>
<td>78.71</td>
<td>75.11*</td>
</tr>
<tr>
<td>Only-Adj</td>
<td>72.25*</td>
<td>74.50</td>
<td>76.63</td>
<td>71.98*</td>
<td>73.93*</td>
</tr>
<tr>
<td>One-Word</td>
<td>80.75</td>
<td>80.00</td>
<td>75.68</td>
<td>79.54</td>
<td>78.23</td>
</tr>
<tr>
<td>Full</td>
<td>80.25</td>
<td>80.00</td>
<td>76.37</td>
<td>80.16</td>
<td>78.74</td>
</tr>
</tbody>
</table>

*Statistically significant using the chi-square test, p < 0.05.

Evaluation: Domain (in)dependence

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>Epinions 1</th>
<th>Epinions 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pos-F</td>
<td>Neg-F</td>
</tr>
<tr>
<td>Books</td>
<td>0.69</td>
<td>0.74</td>
</tr>
<tr>
<td>Cars</td>
<td>0.90</td>
<td>0.89</td>
</tr>
<tr>
<td>Computers</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>Cookware</td>
<td>0.74</td>
<td>0.58</td>
</tr>
<tr>
<td>Hotels</td>
<td>0.76</td>
<td>0.67</td>
</tr>
<tr>
<td>Movies</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>Music</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>Phones</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>Total</td>
<td>0.81</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Manfred Stede (Uni Potsdam)
Comparative Evaluation on Tweets

<table>
<thead>
<tr>
<th>Polarity-Changing Factors</th>
<th>HL</th>
<th>TBD</th>
<th>System Scores</th>
<th>JRK</th>
<th>KLCH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Macro $F_1$</td>
<td>Micro $F_3$</td>
<td>Macro $F_1$</td>
<td>Micro $F_3$</td>
<td>Macro $F_1$</td>
</tr>
<tr>
<td>All</td>
<td>0.615</td>
<td>0.685</td>
<td>0.593</td>
<td>0.671</td>
<td>0.666</td>
</tr>
<tr>
<td>Negation</td>
<td>0.622</td>
<td><strong>0.691</strong></td>
<td>0.596</td>
<td>0.672</td>
<td><strong>0.641</strong></td>
</tr>
<tr>
<td>Intensification</td>
<td>NA</td>
<td>NA</td>
<td>0.595</td>
<td>0.672</td>
<td>NA</td>
</tr>
<tr>
<td>Other Modifiers</td>
<td>NA</td>
<td>NA</td>
<td><strong>0.613</strong></td>
<td><strong>0.684</strong></td>
<td>NA</td>
</tr>
</tbody>
</table>

U. Sidaranka: Sentiment analysis on German Twitter. Forthcoming dissertation, Univ. Potsdam

Overview

- Sentiment analysis: Introduction, Terminology
- One system: SO-CAL
- **Application: Political text**
- Application: Narrative
- Viewpoint: Sentiment / DH
(1) „Tweet sentiment viz“

https://www.csc2.ncsu.edu/faculty/healey/tweet_viztweet_app/

(2) „Semantic visions“

https://semantic-visions.com
(3) 2008 Primaries: Twitter

7000 Tweets / candidate

**Aspect extraction**
- correlation between noun phrases and cand. names via PMI

**Sentiment**
- use prior polarities from existing lexicon (+ domain-specific adjectives); count
- context: directly preceding negations
- no evaluation


(4) ZDF Politbarometer

- **Student project**
- Try **automatic** analysis (manually-extended lexicon + rules) with context analysis => too difficult!
- **Manual** analysis of 8000 Tweets => not easy

(4) ZDF Politbarometer


Interior minister De Maizière on cancelling the football match D - NL (Nov 2015):

„Teile dieser Antworten würden die Bevölkerung verunsichern“
(5) Crisis perception in UNRWA reports

- **United Nations Relief for Palestine Refugees (UNRWA)**
  - financed largely by voluntary contributions
  - resource mobilization is important, especially in times of unexpected demand for action
  - Data set: annual reports 1951-2016

- **Hypothesis**: Under conditions of policy crisis, international bureaucracies are expected to signal increased budgetary stress to principals, donors and/or the public through documents or speech produced by the bureaucracy.


(5) Crisis perception in UNRWA reports

- **Assumption**: budgetary pressure coincides with sentiment: negative polarity

- **Lexicoder Sentiment Dictionary** (Young/Soroka 2012)
  - designed to capture the sentiment of political texts
  - 1710 positive and 2858 negative words

- Compute negativity of report: neg words / all sentiment words

- **Custom-built lexicon for matters of budget**
  - 18 budgetary pressure terms
  - 11 resource mobilization-related terms
  - 9 terms of financial matters, in particular income
  - 4 expenditure-related terms

- Compute share of budgetary pressure terms among all budget terms in report

Manfred Stede (Uni Potsdam)
(5) Authors’ conclusions

- Interpretation has to be done with caution, but:
  - Sentiment detection is able to detect shifts in negativity over time,
  - which can be interpreted as shifts of crisis perception. (Correlation with events)
  - Diverging results in late 1980s: policy crisis increases, budgetary stress decreases

=> due to successful resource mobilization?


(6) Parliamentary debate

- House of Commons (GB)
  - Motion
  - Speeches
  - Vote
- Motion often contains sentiment (polarity)

=> polarity in speeches needs to be interpreted relative to the motion polarity!

- Corpus: 1997 - 2017
  - 1251 motion-speech units from 129 debates
  - unit has max. 5 utterances (avg. 1049 words)
  - motion also gets a government/opposition label
  - speech gets both a text-based label and a vote-based label
  (concurrency: 92.8%)
Overview

• Sentiment analysis: Introduction, Terminology
• One system: SO-CAL
• Application: Political text
• Application: Narrative
• Viewpoint: Sentiment / DH

(1) Sentiment in novels

Fotis Jannidis: Bedeutungsanalyse und distant reading. Ringvorlesung „Digital Humanities“ (BBAW), 16.1.2018
(2) Sentiment in Lessing’s dramas

- **Manual annotation** of 200 snippets
  - agreement: kappa 0.4 => quite difficult
- **Automatic analysis** using different German sentiment dictionaries
  - accuracy: 70% (much lower than in other tasks)
- **Problems**
  - Orthography (e.g., betriegen, bös)
  - Vocabulary (e.g., verdrießlich, pfui)
  - Diachronic change (e.g., Freier)

(3) Character constellations

Overview

- Sentiment analysis: Introduction, Terminology
- One system: SO-CAL
- Application: Political text
- Application: Narrative
- **Viewpoint: Sentiment / DH**
Summary

Methods
• Lexicon-based vs machine learning
• Word counting
• Graded judgement
• Context: negation, irrealis, ...
• Example: SO-CAL
• Polar facts
• Aspect-based analysis
• Fine grain: source, target, opinion
• Entity level: sentiment „flow”

Applications
• Product reviews
• Political text
  – Word matching for Twitter analysis
  – Word matching for crisis perception in annual reports
  – Parliamentary debate
• Narrative
  – Polarity development in drama
  – Happy end?
  – Character constellations

Viewpoint: Sentiment analysis in DH

• Sentiment analysis has great potential for interesting and fruitful applications in the Humanities.
Viewpoint: Sentiment analysis in DH

• Using quite simple methods, for certain texts and purposes one can achieve good results.
• But not for all texts and purposes.
• The „wordle trap“
  – quality evaluation?
  – dependence on domain and genre
  • (e.g., small)

Manfred Stede (Uni Potsdam)
Viewpoint: Sentiment analysis in DH

• „Read close (too)!“
• Context matters – but in what ways, exactly?
• Domain and genre matter – but where do we draw the lines?
• Annotation guidelines for difficult sentiment problems => let’s debate!
• Annotated corpora => let’s debate!
• => Machine learning

Thank you!