


# Opinion Mining

DFKI, LT-Lab

---

---

LT1



## Outline

- 👉 Introduction
  - Definition of subjectivity and opinion
  - Opinion mining as a language technology
- 👉 Research areas of opinion mining
- 👉 Dropping Knowledge Project
- 👉 Summarization

---

---

LT1

## Subjectivity



☞ “Subjective expressions are words and phrases being used to express opinions, emotions, evaluations, speculations, etc.” (Wiebe et al., 2005).

☞ A general covering term for the above cases is private state:

“a state that is not open to objective observation or verification” (Quirk et al., 1985)

---



---

LT1

## Three main types of subjective expressions (Wiebe & Mihalcea, 2006)



☞ references to private states

- He **absorbed** the information quickly.
- He was **boiling with anger**.

☞ references to speech (or writing) events expressing private states

- UCC/Disciples leaders **roundly condemned** the Iranian President's **verbal assault** on Israel.
- The editors of the left-leaning paper **attacked** the new House Speaker.

☞ expressive subjective elements

- That doctor is a **quack**.

---



---

LT1

## Opinion (Wikipedia)



- ✎ In general, an opinion is a subjective belief, and is the result of emotion or interpretation of facts.
- ✎ An opinion may be supported by an argument, although people may draw opposing opinions from the same set of facts.
- ✎ In casual use, the term “opinion” may be the result of a person's perspective, understanding, particular feelings, beliefs, and desires. It may refer to unsubstantiated information, in contrast to knowledge and fact-based beliefs.
- ✎ Collective or professional opinions are defined as meeting a higher standard to substantiate the opinion.

LT1

## Opinion Mining



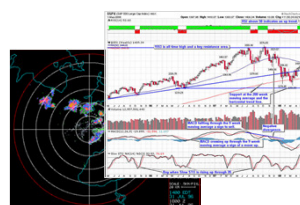
- ✎ Synonym: sentiment analysis
- ✎ Definition:
  - refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials. (Wikipedia)

LT1

## Motivations of Opinion Mining



- ◆ There is a lot of information to discover in online fora and discussions, news reports, client emails or blogs for
  - market research
  - media monitoring and
  - public opinion research
  
- ◆ Opinion mining is a relevant technology to recognize opinions, emotional attitudes about products, services, persons and other topics.



LT1

## Applications (Liu, 2007)



### 👉 Opinion Monitoring

- Consumer opinion summarization  
*E.g. Which groups among our customers are unsatisfied? Why?*
- Public opinion identification and direction  
*E.g. What are the opinions of the Americans about the European style cars?*
- Recommendation  
*E.g. New Beetles is the favorite car of the young ladies.*

### 👉 Opinion retrieval / search

- Opinion-oriented search engine
- Opinion-based question answering  
*E.g. What do Chinese People think about Greek's attitude to work and to EU?*

LT1

## Key Components of Opinions



### Opinion holder (source)

- The person or organization that holds a specific opinion on a particular object/target

### Opinion target

- A product, person, event, organization, topic or even an opinion

### Opinion content

- A view, attitude, or appraisal on an object from an opinion holder.

### Polarity

- Orientations of sentiments expressed in an opinion, e.g., positive, negative or neutral



LT1

## Example



*Former Chancellor Helmut Kohl* attacked *Angela Merkel*  
in an interview with ....

Opinion holder

Target

Polarität



- ✓ subjective sentence
- ✓ opinion holder, target, polarity
- ✓ negative

LT1



## Contextual Valence Shifter

Polanyi & Zaenen (2004)

In 2004 AAAI spring  
Symposium on Attitude



### Simple Lexical Valence (Polanyi & Zaenen, 2004)

- Valence: lexical items or multi-word terms (sentiment words) that communicate with a negative or positive attitude

PART OF SPEECH	Positive Valence	Negative Valence
Verbs	Boost, Embrace, Ensure, Encourage, Delight, Manage, Ease	Conspire, Meddle, Discourage, Fiddle, Fail, Haggle
Nouns	Approval, Benefit, Chance, Approval Benefit, Credit, Favor, Freedom, Hope	Backlash, Backlog, Bankruptcy, Beating, Catastrophe
Adjectives	Attractive, Better, Brave, Bright, Creative, Dynamic,	Annoying, Awry, Arbitrary, Bad, Botched,
Adverbs	Attractively, ...	Annoyingly, ...

*Table 1. Examples of words with non-neutral valence.*

## Contextual Valence Shifter (Polanyi &amp; Zaenen, 2004)

- **Negatives and Intensifiers**
  - *John is successful at tennis* versus *John is never successful at tennis.*
- **Modals**
  - *If Mary were a terrible person, she would be mean to her dogs.*
- **Presuppositional Items**
  - *It is barely sufficient.*
- **Tense**
  - *This was my favorable car.*
- **Collocation**
  - *It looks expensive.* (about appearance)
- **Irony**
  - *The very brilliant organizer failed to solve the problem.*

Discourse based Contextual Valence Shifter (cont.)  
(Polanyi & Zaenen, 2004)

- **Connectors**
  - *Although Boris is brilliant at math, he is a horrible teacher.*

*Although Boris is brilliant at math, he is a horrible teacher.*

Base valence of terms:		Adjusted computation:	
<i>brilliant</i>	+2	<i>(Although) brilliant</i>	0
<i>horrible</i>	<u>-2</u>	<i>horrible</i>	<u>-2</u>
total score:	0	total score:	-2

Table 3. Example of valence adjustment based on discourse connective.

Discourse based Contextual Valence Shifter (cont.)  
(Polanyi & Zaenen, 2004)

- Discourse Structure
  - *John is a **terrific+** athlete. Last week he **walked 25 miles** on Tuesdays. Wednesdays he walked **another 25 miles**. Every weekend he **hikes at least 50 miles** a day.*
- Multi-entity Evaluation
  - ***Coffee is expensive**, but **Tea is cheap**.*
- Comparative
  - *In market capital, Intel is way ahead of AMD.*

OM – Linguistic Resource of OM (Esuli, 2006)

- Linguistic resource of OM are **opinion words or phrases** which are used as instruments for sentiment analysis. It also called polar words, opinion bearing words, subjective element, etc.
- Research word on this topic deal with **three main tasks**:
  - Determining term **orientation**, as in deciding if a given Subjective term has a Positive or a Negative slant
  - Determining term **subjectivity**, as in deciding whether a given term has a Subjective or an Objective (i.e. neutral, or factual) nature.
  - Determining the **strength** of term attitude (either orientation or subjectivity), as in attributing to terms (real-valued) degrees of positivity or negativity.
- Example
  - Positive terms: *good, excellent, best*
  - Negative terms: *bad, wrong, worst*
  - Objective terms: *vertical, yellow, liquid*

## Orientation of terms (Esuli, 2006)

### The problem:

Determining if a **subjective term** has a **Positive** or a **Negative** orientation.

[Hatzivassiloglou and McKeown, 1997]

Hypothesis: *adjectives* in **and** conjunctions usually have similar orientation, though **but** is used with opposite orientation.

### Example (conjunction of adjectives)

- ① The tax proposal was *simple and well received...*
- ② The tax proposal was *simplistic but well received...*
- ③ \* The tax proposal was *simplistic and well received...*

Method: a **weighted graph of similarity of orientation** is defined by analyzing **conjunctions** of adjectives in unprocessed text, then a **minimum-cut** method is applied to the graph.

## Orientation of terms (Esuli, 2006)

[Esuli and Sebastiani, 2005]

Hypothesis: terms with similar orientation have similar glosses.

### Example (glosses for terms with similar orientation)

good: "*that which is pleasing or valuable or useful*";  
"*agreeable or pleasing*".

beautiful: "*aesthetically pleasing*".

pretty: "*pleasing by delicacy or grace; not imposing*".

Each term is represented by its **gloss**.

A binary classifier is learned, in a **semi-supervised** process, using the glosses of the Positive and Negative terms in the training set.

### Orientation of terms (Esuli, 2006)

[Turney and Littman, 2003]

Hypothesis: terms with similar orientation tend to co-occur in documents.

The *Semantic Orientation* (SO) of a term is estimated by combining a **pointwise mutual information** (PMI) measure of the term against some *paradigmatic terms*.

Pos = {good, nice, excellent, positive, fortunate, correct, superior}

Neg = {bad, nasty, poor, negative, unfortunate, wrong, inferior}

PMI is measured using the number of results returned by the AltaVista search engine.

$$PMI(t, t_i) = \log \frac{\#("t \text{ NEAR } t_i")}{\#("t")\#("t_i")}$$

$$SO(t) = \sum_{t_i \in Pos} PMI(t, t_i) - \sum_{t_i \in Neg} PMI(t, t_i)$$

### OM – Polarity acquisition of lexicons

- Application:
  - Naive solution to achieve prior polarities
- Problem:
  - Mixture of subjective & objective words
    - E.g. *long & excellent*
  - Conflict
    - E.g. Nice and Nasty ( the first hit from Google for "Nice and \*")
  - Context dependent
    - E.g. It looks cheap. It is cheap.
    - E.g. It is expensive. It looks expensive.

## OM – Research topics

- Development of linguistic resources for OM
  - Automatically build lexicons of subjective terms
- **At the document/sentence level**
  - Simple opinion extraction (a holder, an object, an opinion)
  - Subjective / objective classification
  - Sentiment classification: positive, negative and neutral
  - \* Less information, more challenges
- At the feature level
  - Identify and extract commented features
  - Determine the sentiments towards these features
  - Group feature synonyms
- Comparative opinion mining
  - Identify comparative sentences
  - Extract comparative relations from these sentences



## OM – Document Level Sentiment Analysis

- Unsupervised review classification
  - Turney, 2003
- Sentiment classification using machine learning methods
  - Pang et al., 2002, Pang and Lee, 2004, Whitelaw et al., 2005
- Review classification by scoring features
  - Dave, Lawrence and Pennock, 2005



### OM – Document-level Sentiment Classification

- Motivation: Determining the overall sentiment properties of a text
- Advantage:
  - Coarse-grained Analysis
  - Detection of a general sentiment trend of a document
- Problem:
  - Different polarities, topics and opinion holders in one document, e.g.

*This film should be brilliant. The characters are appealing. Stallone plays a happy, wonderful man. His sweet wife is beautiful and adores him. He has a fascinating gift for living life fully. It sounds like a great story, however, the film is a failure.*

### Unsupervised review classification

- Hypothesis: the orientation of the whole document is the sum of the orientation of all its parts
- Three steps
  - POS Tagging and Two consecutive word extraction (e.g. JJ NN)
  - Semantic orientation estimation (AltaVista near operator)
    - Pointwise mutual information
 
$$PMI(word_1, word_2) = \log_2 \left( \frac{P(word_1 \wedge word_2)}{P(word_1)P(word_2)} \right)$$
    - Semantic orientation
 
$$SO(\text{phrase}) = PMI(\text{phrase}, \text{"excellent"}) - PMI(\text{phrase}, \text{"poor"})$$
  - Average SO Computation of all phrases
    - The review is recommended if average SO is positive, not recommended otherwise
- The average accuracy on 410 reviews is 74%, ranging from 84% for automobile reviews to 66% for movie reviews

### Others methods

- [Pang et al., 2002]
  - Apply some standard supervised automatic text classification methods to classify orientation of movie reviews
    - Learners: Naive Bayes, MaxEnt, SVM
    - Features: unigrams, bigrams, adjective, POS, position
    - Preprocessing: negation propagation
    - Representation: binary, frequency
  - 82.9% accuracy, on a 10-fold cross validation experiments on 1,400 movie reviews (best from SVM, unigrams, binary)
- [Pang and Lee, 2004]
  - A sentence subjectivity classifier is applied, as preprocessing, to reviews, to filter out Objective sentences.
  - Accuracy on movie reviews classification raises to 86.4%
- [Whitelaw et al. 2005]
  - Appraisal features (i.e. appraisal groups are detected are added to the Movie Review Corpus, which obtained a 90.2% classification accuracy.

### OM – Sentence-level Sentiment Classification

- Advantage:
  - Even though the analysis is still coarse, it is more specific than document-level analysis
  - The results can be reused as input for document-level classification
- Problem:
  - Multiple sentiment expressions with different polarities, e.g.

*The very brilliant organizer failed to solve the problem.*

## OM – Sentence Level Sentiment Analysis (cont.)

- [Riloff and Wiebe, 2003]: subjective / objective classification
  - Taking advantages of Information Extraction techniques
  - Manually collected opinion words + AutoSlog-TS

<subject> passive-vp	<subj> was satisfied
<subject> active-vp	<subj> complained
<subject> active-vp dobj	<subj> dealt blow
<subject> active-vp infinitive	<subj> appears to be
<subject> passive-vp infinitive	<subj> was thought to be
<subject> auxiliary dobj	<subj> has position
active-vp <dobj>	endorsed <dobj>
infinitive <dobj>	to condemn <dobj>
active-vp infinitive <dobj>	get to know <dobj>
passive-vp infinitive <dobj>	was meant to show <dobj>
subject auxiliary <dobj>	fact is <dobj>
passive-vp prep <np>	opinion on <np>
active-vp prep <np>	agrees with <np>
infinitive prep <np>	was worried about <np>
noun prep <np>	to resort to <np>

## OM – Research topics

- Development of linguistic resources for OM
  - Automatically build lexicons of subjective terms
- At the document/sentence level
  - Simple opinion extraction (a holder, an object, an opinion)
  - Subjective / objective classification
  - Sentiment classification: positive, negative and neutral
  - \* **Less information, more challenges**
- At the feature level
  - Identify and extract commented features
  - Group feature synonyms
  - Determine the sentiments towards these features
- Comparative opinion mining
  - Identify comparative sentences
  - Extract comparative relations from these sentences



### OM – Feature-based OM and Summarization (Hu and Liu, 2004)

**GREAT Camera.** Jun 3, 2004  
 Reviewer: jprice174 from Atlanta, Ga.  
 I did a lot of research last year before I bought this camera... It kinda hurt to leave behind my beloved nikon 35mm SLR, but I was going to Italy, and I needed something smaller, and digital.  
 The pictures coming out of this camera are amazing. The 'auto' feature takes great pictures most of the time. And with digital, you're not wasting film if the picture doesn't come out. ...

**Feature1: picture**  
 Positive: 12  
 • The pictures coming out of this camera are amazing.  
 • Overall this is a good camera with a really good picture clarity.  
 ...  
 Negative: 2  
 • The pictures come out hazy if your hands shake even for a moment during the entire process of taking a picture.  
 • Focusing on a display rack about 20 feet away in a brightly lit room during day time, pictures produced by this camera were blurry and in a shade of orange.

**Prior & contextual SO**

- E.g. *Hotel Review:*
  - hot water
  - hot room
- E.g. *Car Review*
  - looks expensive
  - is expensive

**Feature extraction:**

- Explicit & Implicit
  - E.g. great photos <photo>
  - E.g. small to keep <size>
- Frequent & Infrequent

Summary of reviews of Digital camera 1

Comparison of reviews of Digital camera 1 and Digital camera 2

### Featured-based – Feature Extraction

- Frequent & Infrequent features
  - Frequent feature: Label sequential rules
    - Annotation
      - "Included memory is stingy"
      - <{included, VB}{feature, NN}{is, VB}{stingy, JJ}>
    - Learned LSRs
      - <{easy, JJ}{to}{\*, VB}> -> <{easy, JJ}{to}{feature, VB}>
    - Feature extraction
      - The word that matches \$feature is extracted
  - Infrequent feature
    - Observation: the same opinion word can be used to describe different features and objects
      - E.g. The pictures (high-freq) are absolutely amazing.
      - E.g. The software (low-freq) that comes with it is amazing.

Frequent features

Infrequent features

Opinion words

### Featured-based – Group Feature Synonyms


- **Identify part-of relationship (Popescu and Etzioni, 2005)**
  - Each noun phrase is given a PMI score with part discriminators (e.g. of scanner, scanner has) associated with the product class, (e.g. a scanner class)
- **Carenini et al., 2005 is based on similarity metrics**
  - The system merges each discovered feature to a feature node in the pre-set taxonomy
  - The similarity metrics are defined based on string similarity, synonyms and other distances measured using WordNet

Camera	Image
Lens	Image Type
Digital Zoom	TIFF
Optical Zoom	JPEG
...	...
Editing/Viewing	Resolution
Viewfinder	Effective Pixels
...	Aspect Ratio
Flash	...
...	...

### Feature Extraction and Group

- Advantage:
  - Precise sentiment analysis about explicit features
- Problems:
  - Multiple relations
    - *Gas Mileage of VW Golf is great.*
      - Entity: VW Golf
      - Attribute: Gas Mileage
  - Domain knowledge intensive:
    - *V12 8000CC is pretty powerful.* <automobile engine version>
    - *V6 4000CC is not a real good engine.*
- WordNet is too general

## OM – Research topics

- Development of linguistic resources for OM
  - Automatically build lexicons of subjective terms
- At the document/sentence level
  - Assumption: each document, sentence or clause focuses on a single object and contains opinion (positive, negative and neutral) from a single opinion holder
  - Subjective / objective classification
  - Sentiment classification: positive, negative and neutral
  - \* **Less information, more challenges**
- At the feature level
  - Identify and extract commented features
  - Group feature synonyms
  - Determine the sentiments towards these features
-  **Comparative opinion mining**
  - Identify comparative sentences
  - Extract comparative relations from these sentences



## Featured-based Sentiment Orientation (Popescu and Etzioni, 2005)

- **Contextual Semantic Orientation**
  - $\langle \text{word}, \text{SO} \rangle$ ,  $\langle \text{word}, \text{feature}, \text{SO} \rangle$ ,  $\langle \text{word}, \text{feature}, \text{sentence}, \text{SO} \rangle$ 
    - E.g. S1: "I am not happy with this sluggish driver."
      - $\langle \text{sluggish}, ? \rangle$ ,  $\langle \text{sluggish}, \text{driver}, ? \rangle$ ,  $\langle \text{sluggish}, \text{driver}, \text{S1}, ? \rangle$
- **Relaxation labeling: sentiment assignment to words satisfying local constraints.**
  - *Constraints:*
    - conjunctions, disjunctions, syntactic dependency rule, morphological relationships, WordNet-supplied synonymy and antonymy, etc.
  - *Neighborhood:* a set of words connected the word through constraints.
    - E.g. "hot(?) room and broken(-) fan" → hot(-)

**dropping knowledge**

ask, answer and discuss questions, download media, and use the searchable database to find social topics that matter most to you. **the global dialog platform**

SEARCH VISUAL BROWSE ASK A QUESTION HOME | BLOG | ABOUT US | DONATE NOW! | HELP

← go back Overview

**Ethnicity & Culture**

- Sociosphere
- Melting pot
- Norms
- Ethnocide
- Cultural identity
- Nationalism
- Humanities
- Minorities
- Ethnocentrism
- Multiculturalism
- Indigenous people
- Biculturalism
- Tribalism
- Assimilation
- Belief
- Diffusion
- Traditions
- Habits
- Polygamy
- Peaceful co-existence
- Values
- Bilingualism

**dropping knowledge**

ask, answer and discuss questions, download media, and use the searchable database to find social topics that matter most to you. **the global dialog platform**

SEARCH VISUAL BROWSE ASK A QUESTION HOME | BLOG | ABOUT US | DONATE NOW! | HELP

Please enter your query quick search Open advanced search or Add search criteria

**Most active media**

- the table of free voices, answers (11,200 video answers)
- commercial gallery (they asked themselves ...)
- Why is it socially acceptable to hoard [...] by Question Donor 9046
- Technically, God doesn't have religion. You... by Elisabet Sahtouris
- It is a simple thing. It is a question of [...] by Gladman Chibememe
- Let's start with the true story of [...] by Sima Wali

**Hot topics**

animal rights arts & literature children & youth issues **democracy & freedom**



## References

- Slides
  - <http://medialab.di.unipi.it/web/Language+Intelligence/OpinionMining06-06.pdf>
  - <http://www.cs.uic.edu/~liub/opinion-mining-and-search.pdf>
  - <http://www.cs.cornell.edu/home/lee/talks/lee-aaai08.pdf>
- Papers, books or chapters
  - Bo Pang and Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval*, Vol. 2, <http://www.cs.cornell.edu/home/lee/omsa/omsa-published.pdf>
  - Bing Liu. "Sentiment Analysis and Subjectivity." *Invited Chapter for the Handbook of Natural Language Processing, Second Edition*, March, 2010. <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>
  - Hatzivassiloglou, Vasileios and Kathy McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97)*, pages 174–181, Madrid, Spain.
  - Hu, Mingqiang and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2004 (KDD 2004)*, pages 168–177, Seattle, Washington.
  - A. Popescu, "Extracting Product Features and Opinions from Reviews", *Oren Etzioni, Proceedings of HLT-EMNLP, 2005*
  - Wilson, Theresa, Janyce Wiebe, and Paul Hoffman. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pages 347–354, Vancouver, Canada.
  - X. Cheng, *OMINE: Automatic Topic Term Detection and Sentiment Classification for Opinion Mining*, Master thesis, 2007
  - ...