

Opinion Mining

Exercises

Feiyu Xu

DFKI



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)



What are the components of an Opinion

- 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

What are the subtasks of Opinion Mining

- Subjectivity classification
- Polarity classification
- Opinion extraction

Opinion Mining – Research topics

- Development of linguistic resources for opinion mining
 - 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
- 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



Please show at least two methods for acquisition of sentiment word lexicons

- conjunctions
- semantic orientation via PMI
- similar glossary description

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 – 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.***