

Sentiment Analysis using Word Sense Disambiguation Techniques

Vassiliki Rentoumi

IIT, Software and Knowledge Engineering Laboratory (SKEL) NCSR
'Demokritos', Athens, Greece

Table of Contents

Introduction

Research areas

State of the Art

Challenges

Our approach

Sentiment/Subjectivity: Definitions

What do we mean by subjectivity:

- ▶ The linguistic expression of somebody's emotions, sentiments, evaluations, opinions, beliefs, speculations, generally of private states.
- ▶ Private states are those states which are not open to objective evaluation or verification (Quirk, Greenbaum, Leech, Svartvik (1985))
- ▶ Private states are expressed in language using three ways:
 - ▶ Most American people **believe** that Bush is a bad President.(direct subjective expression)
 - ▶ The editors of the left-leaning paper **attacked** the new House Speaker.(private states expressed in speech event)
 - ▶ The part the US human rights report about China is **full of absurdities and fabrications**(expressive subjective element)

Sentiment/Subjectivity:Definitions

What do we mean by sentiment:

- ▶ Sentiment Analysis is a subset of Subjectivity Analysis . Sentiment is the type of subjectivity, specifically involving positive or negative opinion, emotion or evaluation.
- ▶ Is the expression of negative or positive state towards a topic.
 - ▶ Many people **liked** this theatrical play although critics **buried** it.

Sub-Tasks of Sentiment Analysis

- ▶ Determine Subjective-Objective polarity
Is the text or language factual or an expression of an opinion?
- ▶ Determine Positive-Negative polarity
Does the subjective text express a positive or negative opinion of the subject matter?
- ▶ Determine the strength of the opinion
Is the opinion weakly positive/negative, strongly positive/negative.

Sentiment Analysis: Research areas

Automatic Sentiment Analysis falls into two main research areas:

- ▶ Determine words or phrases that are associated with subjectivity (private states and sentiments)
e.g believe = private statement, good = positive sentiment.
- ▶ Classification of sentences, clauses, phrases, or word instances within context. e.g either subjective/objective classifications or positive/negative sentiment classification, or combining subjectivity as well as sentiment classification. (Subjectivity analysis is considered as the intermediate step, for sentiment analysis, Esuli, Sebastiani 2006)

Sentiment Analysis: Research areas

- ▶ Exploits automatic sentiment analysis in applications such as:
 - ▶ Product review mining
e.g. What features of the ThinkPad T43 do customers like and which do they dislike?
 - ▶ Review classification
e.g. Is a review positive or negative toward the movie?
 - ▶ Opinion Question Answering Systems
e.g. What is the international reaction to the reelection of Robert Mugabe as President of Zimbabwe?
African observers generally approved of his victory while western Governments denounced it
 - ▶ Tracking sentiments toward topics over time
e.g. Is anger ratcheting up or cooling down?
 - ▶ E-learning environments.

Other Work

- ▶ Assignment of fine-grained affect labels (emotion labeling) based on various psychological theories (Valitutti et al., 2004; Strapparava and Mihalcea, 2007)
- ▶ Detection of:
 - ▶ opinion holders (Kim and Hovy, 2004; Kim and Hovy, 2005; Kim and Hovy, 2006; Choi et al., 2005; Bethard et al., 2004; Kobayashi et al., 2007)
 - ▶ opinion targets (Hurst and Nigam, 2004; Gamon and Aue, 2005; Hu and Liu, 2004; Popescu and Etzioni, 2005; Kim and Hovy, 2006; Kobayashi et al., 2007)
 - ▶ perspective (Lin et al., 2006)

Other Work

- ▶ pros and cons in reviews (Kim and Hovy, 2006a)
- ▶ bloggers' mood (Mishne and Glance, 2006; Mishne, 2005; Leshed and Kaye, 2006)
- ▶ happiness (Mihalcea and Liu, 2006)
- ▶ politeness (Roman et al., 2005)
- ▶ Assignment of ratings to movie reviews (Pang and Lee, 2005)
- ▶ Prediction of election results (Kim and Hovy, 2007)

Methods used for sentiment analysis

Previous work on sentiment analysis focused mainly on determining terms denoting subjectivity or terms denoting positive or negative orientation.

- ▶ The aim was to find relevant words, phrases, or patterns that can be used to express subjectivity or polarity.
- ▶ Determine the polarity of subjective words or phrases.

These methods are still used in the most of the state of the art methods for sentiment analysis combined with other techniques.

Determining words or phrases expressing polarity

Determining the orientation of words : mainly adjectives, verbs, nouns or phrases.(Hatzivassiloglou and McKeown 1997, Wiebe 2000; Kamps & Marx 2002)

- ▶ determining positive, negative and subjective adjectives
e.g. pos = honest, neg = harmful , subj = odd
- ▶ determining positive, negative, subjective verbs and nouns
e.g. verbs: pos = praise neg = blame subj = predict
nouns: pos = pleasure neg = pain subj = feeling
- ▶ determining positive, negative, subjective phrases containing adjectives and adverbs
e.g. pos = excellent efficiency, low cost,
neg = little variation, many troubles

Determining term orientation

How do we identify polar/subjective items?

Based on the assumption that contexts are coherent.

- ▶ Using the conjunction method. Conjunctions, correlate senses with equal or opposite orientation.
e.g. nice and comfortable, beautiful but stupid.
(Hatzivassiloglou & Mc Kowen 1997)
- ▶ Statistical Association: Based on the assumption that if words with the same orientation like to co-occur together, then the presence of one make the other more probable.

Determining term orientation

- ▶ Statistical measures to capture this correlation:
 - ▶ Mutual Information: this measure aims capturing the interdependence between two words. e.g 'unpredictable steering': quering Altavista, asking for this tuple close to 'excellent' and close to 'poor'. The most frequent co-occurrence gives us the orientation of the tuple. (Turney and Littman 2003)
 - ▶ Distributional similarity process: words that have similar senses tend to appear in similar contexts.(Lin 1998,)

Methods used for sentiment analysis

e.g a. slave = (a) bound in servitude, (b) one who is abjectly subservient to a specified person or influence and (c) one who works very hard

b.DSW list workhorse, servant, captive

c. Use wns to find similarity scores for a sense with each of the words in DSW list.

d.Assign subjectivity : Use MPQA corpus annotated for subjective expressions.

- ▶ For each DSW word's instance in coprus
 - ▶ If it is in a subjective expression, add the similarity score to subjectivity score.
 - ▶ else subtract the similarity score from subjectivity score
 - ▶ total sim score = total sim score + similarity score
- ▶ subjectivity score = subjectivity score/total sim score.

Determining term orientation

- ▶ Assume that alternatives are similarly subjective. e.g synonyms share the same polarity where antonyms have opposite polarity.(Esuli Sebastiani 2005, 2006)
 - ▶ Take advantage of WordNet lexical relations.(Esuli Sebastiani, 2006)
 - ▶ Expand seed lists (of positive and negative words), using WordNet lexical relations(synonymy and antonymy).
 - ▶ Expand seed lists of objective words using WordNet relation of hyponymy.

- ▶ The most broadcast method for annotating words or expressions with either subjectivity or polarity is the bootstrapping method. The output of this method is used in order to train a classifier to label terms with pos/neg or pos/neg/obj, or subj/obj.
 - ▶ the benefit of bootstrapping: we only need a small seed corpus of annotated data with polarity or subjectivity.
 - ▶ train classifiers with few seed labelled data to automatically label sentences with polarity (pos/neg) or subjectivity (subj/obj) from unannotated texts.
 - ▶ check and filter the output, seed again the algorithm with the labelled data and iterate till you annotate all the corpus.

Gaps of current methods

- ▶ Document classification according to sentiment: most of the time the sentiment of the whole text is considered as the sum of the sentiments found in the individual sentences.
 - ▶ Find more advanced metrics to combine predictions of sentences.
- ▶ Polarity versus Subjectivity.
 - ▶ Most approaches aim at the classification between positive versus negative sentiment or between subjective versus objective. Lately there has been focus on combining both : subjectivity(subj vs obj) and sentiment classification(pos vs neg). (Esuli, Sebastiani 2006)

Gaps of current methods

- ▶ For sentiment classification words are assigned with prior polarity, although it isn't reassured that these words participate in subjective context (where opinions are expressed). e.g. 'The protagonist tries to protect her good name' = neutral polarity,
- ▶ The example above despite the presence of 'good', tells us nothing about the author's opinion about the movie.
- ▶ Moreover this example could participate in a negative movie review.
- ▶ Thus, polarity detection (negative or positive oriented phrases) presupposes subjectivity detection (the place where opinion is expressed)

Gaps of current methods II

- ▶ Wordnet lexical relations are not always a good indicator of polarity detection e.g synonyms may have different polarity:
The smell of gas; The aroma of frying onions; Hospital odors;
The scent of pine needles.
 - ▶ The nouns smell and odour express neutral polarity.
Specifically, the noun smell when it is not accompanied by an adjective it bears a negative regard. e.g what's that smell.
 - ▶ Additionally, the nouns aroma and scent are used in a positive oriented environment.
- ▶ Many words have both sentiment-bearing and neutral senses
 - ▶ E.g. great typically tagged as positive, but according to statistics in WordNet, used neutrally 75% of occurrences
 - ▶ Solve this by using sense-tagged word lists.

Exploit senses for sentiment analysis

- ▶ Senses: Maybe we could look for subjectivity and polarity in senses rather than words.
- ▶ subjectivity constitutes a property of language that is correlated with word senses (Michalcea, Word Sense and Subjectivity) it is moreover proved that subjectivity could ameliorate WSD.
- ▶ Which words can potentially participate in subjective contexts?
- ▶ Those that semantically lose their literal senses, and are used with metaphorical and expanded meanings.

Take the challenge

- ▶ It is enough to detect them. How? By disambiguating polysemous words.
- ▶ Finding them (subjective senses) we can moreover detect contextual polarity.
- ▶ Thus we assume that WSD is valuable for subjectivity/sentiment analysis

Sentiment analysis in phrases containing monosemous words

The aforementioned approaches were enough in order to assign sentiment to phrases like:

- ▶ It was his opinion that the patient, a woman, was suffering from tick bite fever.
- ▶ 'The report is full of absurdities Xirosima Sama', said.
In order to handle monosemus words is enough to chose a state-of-the art method that deals with them efficiently.

Sentiment analysis in phrases containing ambiguous words

Its harder to assign sentiment to phrases like:

- ▶ What's the catch? (catch = a hidden drawback)
- ▶ There was a political earthquake. (earthquake = disruptive event)
- ▶ He is such an animal. (animal = brutal person)

because they are ambiguous/polysemous.

We understand that the above phrases bear subjectivity as well as polar orientation.

This happens due to word senses that are activated in the above sentences which are subjective, and thus yielding subjectivity.

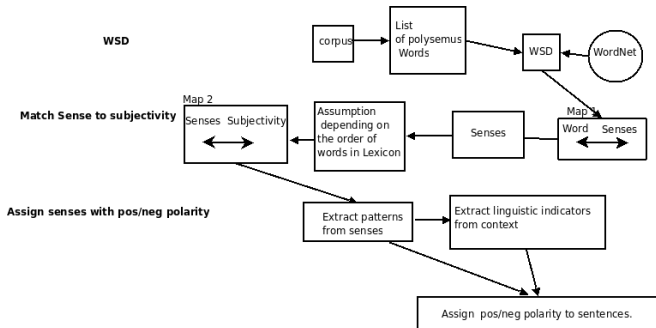
Sentiment analysis in phrases containing ambiguous words: Subjective vs Objective word senses

- ▶ The above mentioned examples use the metaphorical or expanded words senses.
- ▶ Senses, as they are digressing from the literal senses of words, tend to express subjectivity.
- ▶ **Metaphorical and Expanded word senses: Subjective senses**

Polarity could be also detected in the definitions of subjective senses of words.

- ▶ earthquake (something that is severely disruptive; upheaval) It is claimed that selling the company caused an earthquake among the employees
- ▶ dead - (devoid of activity) this is a dead town; nothing ever happens here”
- ▶ boil (be in an agitated emotional state).
He was boiling with anger.

Our Method For Sentiment Analysis



Our Method For Sentiment Analysis

- ▶ Choose a state-of-the art method for WSD.
 - ▶ an approach based on statistical methods that uses linguistic principles.
 - ▶ exploit given linguistic resources (WordNet?), in order to find the proper senses.
 - ▶ using state of the art metrics for measuring semantic similarity and finding the appropriate sense for each word.
 - ▶ the proper sense of the word in the context is derived assuming that a sense of a given word which has most DSW assigned to it is the sense that this word has in the current context.

Our Method For Sentiment Analysis

- ▶ Assign subjectivity scores to word senses, derived from the previous step.
 - ▶ polysemous words once used with their metaphorical or expanded senses tend to be used in subjective expressions.
 - ▶ e.g His alarm grew.
- ▶ Filtering subjective from objective senses.
 - ▶ e.g. The alarm went off.

Our Method For Sentiment Analysis

- ▶ Assign polarity to subjective word senses. Definitions of subjective word senses are very good indicators of polarity orientation as they bear enough semantic and lexical information
 - ▶ alarm, dismay, consternation-(fear resulting from awareness or danger)
- ▶ We are going to exploit these definitions (and glosses) in order to extract the most useful linguistic features that would be good indicators for sentiment analysis

- ▶ Adopt a combined method, linguistic approach with machine learning, that will exploit the sense definitions, as semantic indicators of sentiment.
 - ▶ Annotate few patterns (adverbs-verbs, adjectives-nouns, etc.) from definitions and glosses and use bootstrapping algorithms to annotate all definitions with positive or negative.
 - ▶ Learn patterns from context that denote polarity and combine them with sentiment-tagged senses in order to obtain contextual polarity.

Corpus

MPQA

Opinion corpus

Product Reviews

SemEval-07 data set

Level of Annotation & type(s)

Phrases & sentences & Private states

Expressions and Sentences & Subj, Obj.

Product Features & Sentiment Features

Headlines & Sentiment

Thank you!
Questions...?