

Robust Compositional Polarity Classification

Manfred Klenner & Stefanos Petrakis & Angela Fahrni

Computational Linguistics
Zurich University, Switzerland

{klenner, petrakis}@cl.uzh.ch

angela.fahrni@swissonline.ch

Abstract

We describe a pattern-based system for polarity classification from texts. Our system is currently restricted to the positive, negative or neutral polarity of phrases and sentences. It analyses the input texts with the aid of a polarity lexicon that specifies the prior polarity of words. A chunker is used to determine phrases that are the basis for a compositional treatment of phrase-level polarity assignment. In our current experiments we focus on sentences that are targeted towards persons, be it the writer (I, my, me, ..), the social group including the writer (we, our, ..) or the reader (you, your, ..). We evaluate our system on a manually annotated set of sentences taken from texts from a panel group called 'I battle depression'. We present the results of comparing our system's performance over this gold standard against a baseline system.

Keywords

sentiment analysis, polarity composition

1 Introduction

Polarity classification aims at identifying the positive and negative polarities of text, at various levels, including document, sentence, phrase and word level.

Such a task can be guided by the principle of compositionality[5], which states that:

"The meaning of a complex expression is determined by its structure and the meanings of its constituents."

Based on this principle, the polarity of a portion of text can be composed from the polarities of its constituents in a systematic way[11]. An example of such compositionality appears in the following sentence,

"He is a good liar"

which is classified as negative¹, since a positive adjective and a negative noun yield a negative noun phrase. In principle, such an incremental compositional interpretation might proceed up the sentence level — negating, confirming and intensifying already computed phrase polarities. In the sentence:

"He is a quite good liar."

¹ With an overtone of admiration

the positive polarity of the adjective is confirmed and intensified ('quite'), whereas in the sentence

"He is not an extremely good liar."

the positive polarity of the adjective remains but is decreased. This happens because the adverb 'extremely' is shifted by the "not" negator and does not function as an intensifier anymore. Negation is the most common form of so-called polarity shifters. Another example is 'without' - 'without hope' is negative, but 'without fear' is positive.

In the simplest case, word polarities are provided by a polarity lexicon. Commonly used lexicons are the subjectivity lexicon from [15], the semi-automatically derived SentiWordNet [6] or lexicons generated from the General Inquirer lexicon [12].

Ambiguity turns out to be a problem: 'a cheap therapy' might be regarded as positive if 'cheap' means 'low price' but negative if it means 'low quality'. However, we have identified only few cases of ambiguity in our experiments. In principle, we identify ambiguity of this type as a challenging problem, although we don't cope with it in our current setup.

Another problem is 'out of the blue' non-neutral polarity. That is, combinations of two or more neutral words might yield a non-neutral polarity. For instance, the phrase 'long waiting time (to see the doctor)' is negative, although all parts are neutral. No prior polarity lexicon can cope with these cases. We have proposed a corpus-based approach to solve these cases in [7].

Finally, figurative language such as irony and sarcasm might as well occur in such texts. Consider the following example:

"I also am being charged 100 for missing a doctor's appt. What a way to make me feel better"

The intended meaning of the second sentence clearly is not positive, although the literal interpretation suggests this.

We introduce a system for polarity classification based on the prior lexicon from [15] and the output of the TreeTagger chunker [14].

It is shown that our cascaded, pattern-based compositional polarity determination yields good empirical performance on texts from a self-help group called 'I battle depression'. The evaluation of our system involved constructing a gold standard used for testing a baseline system's performance against our own.

2 Resources and tools

We have searched for texts where people are expressing strong emotions. A website called “the experience project”² has proved interesting for our purposes. On that website, groups can be found to rather diverse topics such as ‘I quit smoking’, ‘I love cats (music, books, lyrics)’, ‘I want to loose weight’ etc. For our experiments, we have taken 2290 texts from a panel group called ‘I battle depression’³.

Here, people explicitly describe their emotional states, their feelings, their experiences, their hopes and fears and even give each other advice how to overcome mental problems such as for instance social anxiety.

In a first step, we wanted to analyse the polarity of phrases and sentences from these texts. In order to achieve this, a polarity lexicon was necessary. We have experimented with the subjectivity lexicon from [15].

The subjectivity lexicon [15] is a resource compiled from various other resources - including the general inquirer (GI). This was done mainly manually, but in part also automatically. The lexicon comprises about 8,000 polarity tagged words (adjectives, verbs, nouns, adverbs), where each word either is positive, negative or objective. A non-objective word also might be weak or strong subjective (we have not used this information).

3 The composition of polarity

The predominant approach in the area of sentiment detection can be characterised as ‘machine learning on top of a bag of word representation of the input data’. There are very few notable exceptions, namely [11] and lately, [3] (see section related work).

The bag of words approach ignores the fact that sentiment interpretation is compositional. To a certain extent, a machine learning algorithm is able to approximate composition, e.g. the effect of negation (‘I don’t like ..’). However, sentiment composition is a phenomenon that can be fixed with a relatively small set of simple rules with very few exceptions. So there is no need to learn these regularities.

ADJ	NOUN	→	NP	Example
NEG	POS	→	NEG	disappointed hope
NEG	NEG	→	NEG	a horrible liar
POS	POS	→	POS	a good friend
POS	NEG	→	NEG	a perfect misery
POS	NEU	→	POS	a perfect meal
NEG	NEU	→	NEG	a horrible meal

Fig. 1: NP composition

Fig. 1 gives the regularities for NP level composition, where an adjective is combined with a noun. The sentiment orientation of the words comes from a pre-compiled polarity lexicon. So for example, the positive adjective ‘perfect’ combined with the negative noun ‘misery’ yields a negative noun phrase.

² Their slogan is: “Share your experiences anonymously. Meet new friends who understand you”.

³ http://www.experienceproject.com/group_stories.php?g=109

Adverbs act as intensifiers, that is, they leave the orientation, but alter the strength. So a ‘very good friend’ is more than just a ‘good friend’ etc.

NP	Prep	NP	→	PP	
POS	to	NEG	→	POS	solution to my problem
POS	for	POS	→	POS	hope for relief
NEG	of	NEG	→	NEG	pain of disappointment
NEG	of	POS	→	NEG	lost of hope

Fig. 2: NP-PP composition

Fig. 2 shows some regularities holding for NP-PP composition. With NP-PP composition, the effect also depends on the preposition.

Verbs might as well bear a polarity orientation. The Verb ‘love’ is positive, ‘hate’ is negative. ‘To enjoy’, ‘to like’, but also ‘to detest’, ‘to dislike’ etc. are all verbs with a clear polarity. The question is, how the combination with their direct objects must be interpreted in terms of compositionality. Is the verbal phrase from the sentence ‘He loves nasty films’ positive or negative, given that ‘nasty films’ is negative. Accordingly, is ‘He hates good books’ positive or negative?

If the mental state of the subject is in question, then the verb overwrites the NP polarity, i.e. the VP with love is positive independent from the polarity of the direct object (accordingly for ‘hate’). If however, the character (in the sense of morality) of the subject is in question, then the VP with love is negative. To love negative things is negative⁴.

Some verbs like ‘fail’, ‘stop’ etc., are polarity shifters. A polarity shifter inverts the polarity of the embedded phrase. ‘Fail to make someone angry’ then is positive: a negative embedded phrase is inverted.

Other polarity shifters are adverbs such as hardly (‘this is hardly true’) and negation (‘I don’t like action films’).

We have implemented our sentiment composition as a cascade of transducers operating on the prior polarities of the subjectivity lexicon, the output of the Tree-Tagger chunker [14] and manually written pattern-matching rules.

4 Cascaded sentiment composition

We propose an engineering approach to sentiment composition, a system combining both domain-specific and domain-independent knowledge. Our system operates based on the assumption that since sentiment composition takes place in a rather canonical and straightforward way and therefore its regularities can be captured by a limited number of rules. This set of rules is what we mentioned as domain-independent knowledge as it takes effect across different domains. The other basic module of this system is the polarity lexicon, and that is - at least in part - domain-specific.

⁴ How to evaluate the following sentence from the group ‘I Quit Smoking’: I like smoking?

It can be modified or completely replaced to suit specific domains. An immediate advantage of such an approach compared to other dominant approaches in the field, namely machine learning, is that we do not need neither a training corpus nor a training phase to bootstrap our system. For our system to operate in a new domain we need only adapt the polarity lexicon. We discuss problems with the determination of the polarity of words in section 6.

Our system receives as input text that has been syntactically chunked. In our current setup we have used the tree-tagger chunker [14] which is currently available for three languages (English, French, German), but other chunkers should also work well with our system.

The chunked text that is inputted to our system is a flat structure, which is evaluated via a cascade of transducers. Simpler rules are taking effect first, and their output is then consumed by more complex ones, moving from word level to sentence level sentiment composition. The rules are written in our own pattern-matching language, devised to facilitate the engineering process. A sample of a basic set of rules is the following:

```
vc_to=_,_POS;nc=_ ->POS
vc_to=_,_NEG;nc=_ ->NEG
?vc_*=_:SHI,*=vb:POS;POS ->NEG
?vc_*=_:SHI,*=vb:POS;NEG ->POS
```

Rule 1 and 2 operate in a similar fashion. A verb chunk (vc) that contains a "to" item and a positive (POS) or a negative (NEG) one, is adjoined with a noun chunk (nc) to compose a positive constituent, e.g. 'to enjoy the sun', or a negative one, e.g. 'to envy the success'. Rules 3 and 4 are also alike, they bring together a shifted (SHI) verb chunk that contains a positive verb (vb:POS) followed by a POS or NEG item to produce a NEG constituent, e.g. 'not succeed to love', or a POS constituent 'not earn the contempt', respectively.

Given for instance the sentence 'I did not achieve to cheer him', we get the chunked text that is shown in Fig. 3.

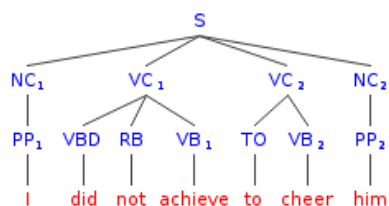


Fig. 3: Example 1

The rules mentioned above would then be applied in the following cascade⁵:

```
to cheer+ him → POS1
I did not* achieve+ POS1 → NEG2
```

⁵ Indices indicate succession, → means 'rewrite' and the polarity of lexical items is indicated by the superscript where '+' means positive, '-' means negative and '*' indicates a polarity shifter

The result is a negative polarity at the sentence level. Another example would be the sentence 'She did not manage to hurt his feelings', seen in Fig. 4.

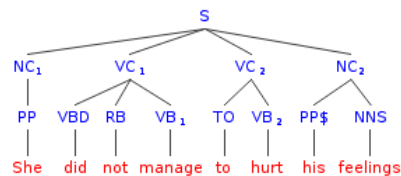


Fig. 4: Example 2

resulting in a positive polarity at the sentence level, evaluated in the following way:

```
to hurt- his feelings → NEG1
She did not* manage+ NEG1 → POS2
```

Another part of our system is polarity strength. Each word has a polarity strength that ranges from 0 to 1. A word with positive polarity and strength 1 is strongly positive, and a negative word with strength 1 is strongly negative. Intensifiers have no polarity but a strength value. Polarity strength adds up while rules are applied, except for intensifiers which are multiplied with word or phrase strength.

For example, 'good friend' yields a positive NP polarity, the polarity strength is the sum of the polarities of 'good' and 'friend' (currently 1 respectively). Intensifiers duplicate the polarity without altering it. So 'a very good friend' has a polarity strength of 4. Shifters such as 'not' invert the polarity without altering the strength. In order to determine sentence-level polarity (e.g. in sentences with more than one target) all phrase-level polarities are added up and the polarity class with the highest strength is chosen (e.g. a sentence has positive polarity, if the sum of positive strength is higher than the sum of negative strength).

5 Empirical evaluation

We have previously [9][10] evaluated our system using customer reviews data as described in [4] and texts from the depression group of the experience project [1]. In those evaluations we have used the same set of composition rules and the same lexicon, the only difference being the selection of targets. In the customer reviews data evaluation, the targets were already identified in the gold standard, while in the depression group texts we set as targets the first person singular, second person singular and first person plural personal pronouns (we call these targets I-targets).

Our system produced promising results during both of these evaluations. We present here yet another type of evaluation for our system in an effort to test it exhaustively. In this evaluation we produce a gold standard from the depression group texts. Then we setup a baseline system and we compare the performance of our system with it. An additional difference from the previous evaluations, is that we focus on sentence rather than on phrase level polarity of the depression group texts.

5.1 Gold standard

We set about building a gold standard, annotated by two different annotators. We work with texts from the depression group of the experience project. From these texts we chose those sentences that contain an I-target. We have a set of 346 sentences, where each sentence is labelled as positive, negative or neutral.

The interannotator agreement was measured first as a simple percentage, which gave us a 68%. We also calculated two more measures of agreement, following [2]. The expected interannotator agreement was 46.5% and finally the chance corrected interannotator agreement was 41.15%. Finally, the set of sentences that both annotators agreed on was selected to be used as our gold standard. That gave us a set of 222 sentences that we used in our evaluation.

5.2 Systems specifications

Our system, PolArt, uses as a prior lexicon the subjectivity lexicon from [15]. We enhanced the lexicon by adding 'not' as well as a few other polarity shifters. We have also added polarity strengths, but we did it uniformly (strength of 1). Only selected words are given a fine-grained polarity strength - in order to carry out some experiments. The set of rules, 70 in total remains the same as the one we used in prior evaluations.

The baseline procedure determines the majority class for each sentence, by examining each word inside the sentence. To examine each word and retrieve a polarity value for it, the baseline system is also using the subjectivity lexicon. The majority class is determined by counting positive and negative words inside the sentence. The most frequent polarity is assigned to the whole sentence. Note, that although the baseline system is not meant to work in a compositional-way, we make an exception for the sake of reliability of the results. This exception covers the special case of a shifter like "not" which inverts the polarity (e.g. 'not guilty' is positive).

5.3 Results

We ran both systems with our gold standard as input. The baseline system correctly classified 64 out of the 222 sentences, which translates into an accuracy of 28.82%. PolArt correctly classified 114 sentences, scoring a 51.35% accuracy which is a rather mediocre score. The interannotator agreement metrics for our gold standard indicate that even for human experts, sentiment classification is a demanding task, neither trivial nor unanimous.

During the evaluation phase, we also asked the annotators to state which of the two systems came closer to their classification of each sentence. We think of this metric as a proximity score⁶. From the 222 sentences of the gold standard, the baseline system got 36.48% on proximity while PolArt got a 63.51%.

In Fig 5, the scores for precision, recall and F-measure are given for negative (NEG), positive (POS) and neutral (NEU) classification of sentences, for each of the two systems.

⁶ This score is more meaningful for these sentences where both of the systems disagreed with the gold standard

[Baseline/Polart]	Recall	Precision	F-measure
NEG	0.20/0.59	0.79/0.74	0.33/0.66
POS	0.48/0.4	0.15/0.19	0.23/0.25
NEU	0.56/0.13	0.16/0.11	0.25/0.12

Fig. 5: Recall, precision and F-Measure scores for the baseline system and the PolArt

PolArt performed rather well when it came to negative classifications and moderately well in the case of positive classifications. In neutral classifications it performed poorly, as did the baseline system.

The explanation for this is that neutral polarity, in both the baseline system and PolArt, occurs as a mutual neutralization of accumulated POS and NEG values. This treatment of neutrality is brittle as it is based on a weak concept. It also stands quite separate from the human perception of neutral polarity, as this was made obvious by the annotated gold standard. The following examples present sentences that the annotators classified as neutral, giving a good idea of their view of where neutral polarity lies:

I 'm not sure if he is a deep enough human being to understand

That 's not her real name , but it will suffice

I may have been melancholy or depressed , but I had no real physical symptoms

I read somewhere that age will increase the hormonal mood problems

In these examples both annotators agreed on the neutral polarity of the sentences, while both of the systems disagreed with the annotators, classifying the sentences either as negative or positive.

In order to improve our system's performance, it is crucial to find another way to define and handle neutrality.

6 Open problems with polarity determination

There are remaining problems with polarity determination to be dealt with in subsequent work:

- composition principles are debatable (or application dependent): 'a perfect⁺ spy⁻' - positive or negative?
- composition principles are not deterministic: if 'a perfect⁺ spy⁻' is positive why then is 'a perfect⁺ hassle⁻' in any case negative?
- words without a prior polarity combine to a non-neutral phrase polarity: 'a cold answer' is negative although both words are neutral
- implicit attitudes and figurative language (irony, even slang): 'I was happy that my stepfather disappeared'. The negative attitude towards the stepfather is only implicitly given.

- disambiguation might be necessary before polarity determination 'a cheap therapy' might be regarded as positive if 'cheap' means 'low price' but negative if it means 'low quality'.

7 Related work

Only a limited number of approaches in the field of sentiment analysis copes with the problem of sentiment composition.

The first, fully compositional account to sentence-level sentiment interpretation on the basis of a manually written grammar is presented in [11]. Since based on a normative grammar, their approach is brittle, while our pattern-matching approach operates well in the presence of noise.

More recently, [3] have introduced a machine learning approach to sentiment composition, but they also have experimented with a pattern-matching approach. Their empirical results are based on the MPQA corpus [15]. In the near future, we shall also experiment with the MPQA corpus to enable a direct comparison.

8 Conclusion and future work

We presented in this paper an engineering approach to dealing with polarity classification. The goal was to perform this task outside of the machine learning paradigm. What we managed to prove is that employing a set of compositional rules and a polarity lexicon can be a feasible solution.

What usually dictates the use of machine learning techniques is domain independence. In the case of sentiment composition, pattern matching rules can operate in a domain independent way. We work with a set of 70 such rules that operate in cascades of rewrite operations. A polarity lexicon is necessary and - although at least partially domain dependent - a moderate sized one like the one we use in our system is not a costly resource.

We worked with texts from a panel group called "I battle depression". We prepared a gold standard, as well as a baseline system to measure our system's performance. The results were encouraging, although there exist various tough points to overcome. Among these is the conceptual and practical treatment of neutral polarity.

The texts we worked with were in English. We have tested our system with German [8] as well as with French[13] texts and have made a demo version⁷ available. We are also experimenting with the use of a dependency parser instead of a chunker, since word order in languages such as French and German is less restricted.

All in all, a pattern-based approach to sentiment analysis seems to be a choice of reason. The polarity lexicon - being the most volatile of our resources and the most dependent on domain specific knowledge - is, however, a good candidate for machine learning.

⁷ Visit <http://www.cl.uzh.ch/kitt/polart/> for a demo version of our system for English, German and French

Acknowledgments

This work is funded by the Swiss National Science Foundation (grant 100015_122546/1).

References

- [1] Experience project, 2009. <http://www.experienceproject.com>.
- [2] R. Artstein and M. Poesio. Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555–596, 2008.
- [3] Y. Choi and C. Cardie. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proc. of EMNLP*, 2008.
- [4] X. Ding and B. Liu. The utility of linguistic rules in opinion mining. In *SIGIR*, 2007.
- [5] D. R. Dowty, R. E. Wall, and S. Peters. *Introduction to Montague Semantics*. Reidel, Dordrecht, 1981.
- [6] A. Esuli and F. Sebastiani. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proc. of LREC-06*, Genova, IT, 2006.
- [7] A. Fahrni and M. Klenner. Old Wine or Warm Beer: Target-Specific Sentiment Analysis of Adjectives. In *Proc. of the Symposium on Affective Language in Human and Machine, AISB 2008 Convention, 1st-2nd April 2008. University of Aberdeen, Aberdeen, Scotland*, pages 60 – 63, 2008.
- [8] M. Klenner. Süsse Beklommenheit und schmerzvolle Ekstase. Automatische Sentimentanalyse in den Werken von Eduard von Keyserling. In *Tagungsband der GSCL- Tagung, Gesellschaft für Sprachtechnologie und Computerlinguistik (GSCL)(erscheint)*, Potsdam, 30.9. - 2.10 2009.
- [9] M. Klenner, A. Fahrni, and S. Petrakis. An Experimental Tool for Polarity Classification of Human Affect from Panel Group Texts. In *Acti 2009, Amsterdam, The Netherlands*, 2009.
- [10] M. Klenner, A. Fahrni, and S. Petrakis. PolArt: A Robust Tool for Sentiment Analysis. In *NODALIDA 2009, Odense, Dänemark, 14-16.5., 2009*.
- [11] K. Moilanen and S. Pulman. Sentiment composition. In *Proc. of RANLP-2007*, pages 378–382, Borovets, Bulgaria, September 27-29 2007.
- [12] M. S. P. Stone, D. Dumphy and D. M. Ogilvie. *The General Inquirer: a computer approach to content analysis*. MIT Press, Cambridge, MA, 1966.
- [13] S. Petrakis, M. Klenner, E. Ailloud, and A. Fahrni. Composition multilingue de sentiments. In *TALN (Traitement Automatique des Langues Naturelles) (Demo paper)*, Senlis, France, 2009.
- [14] H. Schmid. Probabilistic part-of-speech tagging using decision trees. In *Proc. of Intern. Conf. on New Methods in Language Processing*, 1994.
- [15] J. W. T. Wilson and P.Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. of HLT/EMNLP 2005*, Vancouver, CA, 2005.