PolArt: A Robust Tool for Sentiment Analysis

Manfred Klenner & Angela Fahrni & Stefanos Petrakis
Computational Linguistics
Zurich University, Switzerland
{klenner, petrakis}@cl.uzh.ch
angela.fahrni@swissonline.ch

Abstract

We introduce PolArt, a robust tool for sentiment analysis. PolArt is a pattern-based approach designed to cope with polarity composition. In order to determine the polarity of larger text units, a cascade of rewrite operations is carried out: word polarities are combined to NP, VP and sentence polarities. Moreover, PolArt is able to cope with the target-specific polarity of phrases, where two neutral words combine to a non-neutral phrase. Target detection is done with the Wikipedia category system, but also user defined target hierarchies are allowed. PolArt is based on the TreeTagger chunker output, and is customised for English and German. In this paper we evaluate PolArt's compositional capacity.

1 Introduction

Sentiment detection aims at identifying the positive or negative polarities of portions of a text – words, phrases[1] and sentences. Normally, the evaluation of dedicated objects is focussed on, e.g. persons or products and their features and attributes (e.g. the *appearance* of a person, the *price* of a product). Our system, PolArt, uses the Wikipedia category system to derive these domain specific target concepts (targets are sometimes called 'features' in the literature).

The polarity of larger text units comprising two or more polarity tagged words is compositional (Moilanen and Pulman, 2007). For example, a 'bad joke' is negative, since a negative adjective and a positive noun yield a negative noun phrase. Besides bearing a negative or positive polarity, words can be polarity shifters. Negation is the most common form, the 'not' in 'this is not a bad joke' shifts the negative polarity of 'bad joke' to a positive polarity. In the simplest case, word polarities are given by a polarity lexicon, e.g.(Esuli and Sebastiani, 2006). Of course, ambiguity turns out to be a problem: 'a cheap meal' is positive if 'cheap' means 'low price' but negative if it means 'low quality'. Moreover, there are target-specific polarities that emerge from the combination of two neutral words (e.g. the negative 'warm beer'). No prior polarity lexicon can cope with this problem. Even worse, the same neutral word might take, depending on the target object, both polarities, positive and negative. For example, 'cold burger' is negative, while 'cold beer' is positive. However, 'cold burger' could also be used ironically. We have no means to detect pragmatic usages.

We introduce PolArt, a robust tool for sentiment analysis. PolArt is based on the output of the TreeTagger chunker (Schmid, 1994) and it uses Wikipedia categories for target detection. It has a pattern-based compositional sentiment semantics that is based on lexicons that code the prior polarity of words, but also on a target-specific lexicon induced from a seed lexicon and the analysis of additional texts, cf. (Fahrni and Klenner, 2008). Currently, PolArt is customised for English and German. In this paper, we focus on the evaluation of PolArt's sentiment composition, readers interested in the Wikipedia-based target detection and the induction of the target-specific polarity lexicon are referred to (Fahrni and Klenner, 2008).

2 PolArt as a Tool

PolArt is a tool to detect and visualise how targets are evaluated in texts. Figure 1 depicts PolArts's output for texts taken from the fast food domain. On the left-hand side, the recognised targets such as 'coffee' or 'cheeseburger' are shown together with their polarities in the text. With a click on a polarity value of a target (e.g. 'positive') all phrases evaluating the target in the selected way and their frequency appear on the right upper window. A click on a phrase displays the context in the right bottom window highlighting all phrases interpreted by the tool in different colours. The advantages of this output are twofold. First, it enables an engineer to analyse the effect of an annotation rule in a fast way. Secondly, it allows

[1] We do not cope with collocations, currently.

Kristiina Jokinen and Eckhard Bick (Eds.)
NODALIDA 2009 Conference Proceedings, pp. 235–238

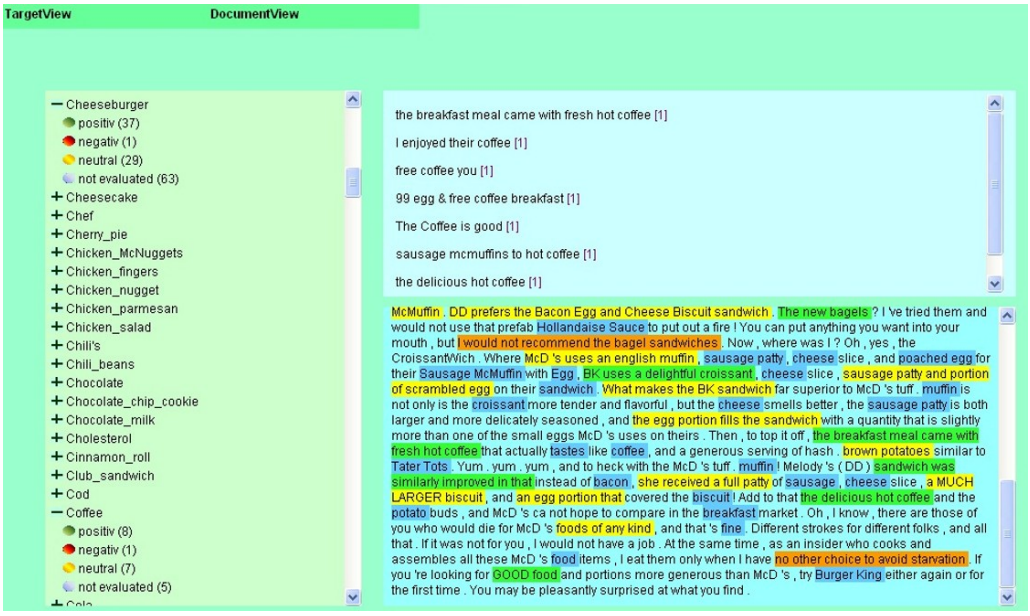

Figure 1: Output of PolArt

users of the tool to quickly get an overview how targets of interest are evaluated in texts and which text passages are important.

3 Target-specific Polarity

As (Turney, 2002) has pointed out, the polarity of some adjectives is domain-dependent. For example, an 'unpredictable plot' clearly increases suspense and as such is positive. An 'unpredictable (behaviour of a) friend' on the other hand is undesirable and thus negative. Please note that the problem with 'unpredictable' has nothing to do with word sense ambiguity (both examples adhere to WordNet word sense 1: not capable of being foretold). Even if the word sense is identified, the polarity still might be open. We call this the *target-specific polarity* of adjectives. An adjective is target-specific, if it takes a polarity dependent on the accompanying noun, e.g. 'old wine' (positive) as compared to 'old bread' (negative). See (Fahrni and Klenner, 2008) for a description and evaluation of that part of our model.

4 Sentiment Composition

The polarity of larger text units comprising two or more words that have a sentiment orientation is compositional. The sentiment orientation of a word comes either from a pre-compiled polarity lexicon or - if it is target specific – has to be learned from domain-specific texts. Available polarity lexicons are e.g. SentiWordNet (Esuli and Sebastiani, 2006) (semi-automatically derived from WordNet) and the subjectivity lexicon introduced in (Wilson et al., 2005). In our experiments, we have used the subjectivity lexicon comprising 8000 words (adjectives, verbs, nouns, adverbs) and their polarities. Our tests with SentiWordNet have been less successful, but see (Fahrni and Klenner, 2008) for an attempt to use a lexicon derived from SentiWordNet.

We have implemented our sentiment composition as a cascade of transducers operating on the prior polarities of the subjectivity lexicon, the output of the TreeTagger chunker (Schmid, 1994) and our pattern-matching rules. The rules for NP level composition comprise the following regularities:

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

At each cascade the matching parts (from the TreeTagger output) are rewritten (simplified). NP rules are applied first, followed by PP rules, verb rules and negation. Given 'He doesn't fail to verify his excellent idea', the cascade is (indices indicate succession, 'excellent' is positive, idea is 'neutral' according to the prior lexicon):

'excellent idea'	→	POS_1
verify POS_1	→	POS_2
fail POS_2	→	NEG_3
not NEG_3	→	POS_4

We have designed a rule language to facilitate the customisation of rules for sentiment composition. Consider these three slightly simplified examples:

```
1. advc_pol=SHIFT;vc_pol=NEG-->POS  % not regret
2. ?nc_no=dt,pol=NEG-->POS          % no problem
3. ?nc_no=dt,pol=POS-->NEG          % no help
```

Rule 1 captures the case where an adverbial chunk (advc) with a polarity shifter (e.g. not) is immediately followed by a verb chunk (vc) with a negative polarity (which has been derived by the application of another rule, or which is simply given by the prior polarity of the verb). The result is a larger chunk with a positive polarity. Rule 2 and 3 are complementary, they capture noun chunks (nc) where a negation (here ' no') precedes a negative or positive word. Again, the polarity is inverted in these cases. Similar rules are designed to determine the polarity of such examples like 'I don't have any complaints' or 'I can't say I like it'. Of course, the flat output structure of a chunker poses limitations on the expressive capacity of such rules. We also have to find ways to evaluate the usefulness of a single rule, i.e an error analysis in terms of false positives and false negatives. We currently work with 70 rules. Since the rules of sentiment composition are domain independent, only the lexicon need to be exchanged (in parts) in order to switch to another domain.

Another, yet experimental part of PolArt, is polarity strength. Each word has a polarity strength that ranges from 0 to 1 (strong positive or negative). Polarity strength adds up while rules are applied. 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. So 'a very good friend' has a polarity strength of 4. Shifter such as 'not' invert the polarity without altering the strength. In order to determine sentence level polarity 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 evaluated PolArt in two steps. The evaluation of the target specific component was done on our own data set (3000 manually annotated noun phrases). The details can be found in (Fahrni and Klenner, 2008). In this paper, we present the evaluation of our composition rules. We have used customer reviews as described in (Ding and Liu, 2007). The authors have manually annotated a number of texts from Amazon[2]. They have identified the targets of the domain (e.g. 'installation software', 'camera') and have numerically qualified their polarity strength (-3 to +3). Here are two examples taken from the dataset:

```
color screen[+2]##it has a nice color screen.
phone[+2],warranty[-2]##this is a very nice
      phone , but there is no warranty on it.
```

In order to generate a gold standard from that data, we have selected those sentences (1511) that contain at least one evaluated target. Gold standard sentence polarity is derived by adding up the polarity strength of all targets of the sentence. If the sum is > 0 then sentence polarity is positive, a zero yields a neutral polarity and a sum of < 0 is negative. For example: 'phone book[+2] speaker-phone[+2]' indicates a positive polarity (since the sum of +2 and +2 is > 0). Cases where two or more targets with inverse polarities are present are, however, problematic. So 'phone[+2],warranty[-2]' indicates a neutral polarity, but 'phone[+2],prize[-1]' would be still positive. In both cases, PolArt would assign a neutral polarity (producing a non-avoidable misclassification), since currently it does not have a full-fledged metric of polarity strength.

The accuracy of the polarity classification at the sentence level in our experiments is 72.46%. Without any rule application, i.e. by just taking the majority class from the sum of the word-level polarities (as a baseline), accuracy is 68.03%. The effect of our compositional component thus amounts to 4.5 %. Unfortunately we can not compare our result with the result of (Ding and Liu, 2007), since these authors have only evaluated their feature extraction component.

Sentence polarity might be regarded as an artificial notion[3], since normally the *targets* appearing in a sentence are getting evaluated. Only in simple cases (sentence with one target) are both viewpoints identical. It is the target-level polarity that is relevant for applications (i.e. which product feature is evaluated 'good', 'poor' etc.). The accu-

[2] cf. www.cs.uic.edu/~liub/FBS/sentiment-analysis.html.

[3] It is, however, an important theoretical problem to determine sentence level polarity.

racy of the polarity classification of the targets is 87.72%. That is: given an evaluated target, PolArt assigns it the right polarity (orientation) in about 9 out of 10 cases. However, 60% of the targets do not receive an evaluation from PolArt, so the accuracy values reported here refer to the found targets (i.e. 40%). The problem here is – among others – that the gold standard data is not very reliable, as some randomly chosen examples suggest. Consider the sentence 'many of our disney movies do not play on this dvd player'. The authors have identified 'disney movie' as a target with a negative evaluation. Neither is true: it is not a target, but if so, it was not negatively evaluated.

As a prior lexicon we have used the subjectivity lexicon from (Wilson et al., 2005). We have added 'not' as a valency shifter and have removed some words that PolArt has identified as target-specific (e.g. low - 'low price' versus 'low quality'). 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.

We have turned off our Wikipedia-based target detection, since the targets are already part of the gold standard information. Note that target detection actually is crucial. For example in the movie domain (another often used domain for sentiment detection), one must well distinguish between content (of a movie) and evaluation. A horror film might get enthusiastic ratings, although the review talks of frightened people, bloodshed and eternal perdition.

6 Related Work

Only a limited number of approaches in the field of sentiment analysis copes with the problem of sentiment composition. A fully compositional account to sentence level sentiment interpretation on the basis of a manually written grammar is presented in (Moilanen and Pulman, 2007). Since based on a normative grammar, their approach is brittle, while our pattern-matching approach operates well in the presence of noise. More recently, (Choi and Cardie, 2008) 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 (Wilson et al., 2005). In the near future, we shall also experiment with the MPQA corpus to enable a more direct comparison (including the pattern-matching part).

7 Conclusion

We have demonstrated that robust sentiment composition with a cascade of polarity rewrite operations and based on a moderate sized polarity lexicon is successful. Our 70 pattern-matching rules are domain-independent, although domain-specific tuning is possible. Domain dependence was one of the main reasons, why pattern-matching approaches have been discarded in the past and have been replaced by machine-learning approaches. This problem is not present in the area of sentiment detection, since polarity composition rules are not specific to the domain of application - only the (target-specific) polarity lexicon is. It has always been acknowledged that carefully designed pattern-based approaches are at least as good as machine learning approaches. A pattern-based approach to sentiment analysis thus seems to be a sensible choice. But domain-specific lexicon induction is a good candidate for machine learning.

Acknowledgement This work is partly funded by the Swiss National Science Foundation (grant 105211-118108).

References

Y. Choi and C. Cardie. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proc. of EMNLP*.

Xiaowen Ding and Bing Liu. 2007. The utility of linguistic rules in opinion mining. In *SIGIR*.

Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A publicly available lexical resource for opinion mining. In *Proc. of LREC-06*, Genova, IT.

A. Fahrni and M. Klenner. 2008. Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Symposion on Affective Language in Human and Machine, AISB Convention*, pages 60 – 63.

Karo Moilanen and Stephen Pulman. 2007. Sentiment composition. In *Proc. of RANLP-2007*, pages 378–382, Borovets, Bulgaria, September 27-29.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proc. of Intern. Conf. on New Methods in Language Processing*.

Peter Turney. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL*, pages 417–424.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. of HLT/EMNLP 2005*, Vancouver, CA.

ISSN 1736-6305 Vol. 4
http://hdl.handle.net/10062/9206