SentiWS – a Publicly Available German-language Resource for Sentiment Analysis

Robert Remus, Uwe Quasthoff, Gerhard Heyer

University of Leipzig, Natural Language Processing Department,
Johannisgasse 26, 04081 Leipzig, Germany
robert.remus@googlemail.com, {quasthoff, heyer}@informatik.uni-leipzig.de

Abstract

SentimentWortschatz, or *SentiWS* for short, is a publicly available German-language resource for sentiment analysis, opinion mining etc. It lists positive and negative sentiment bearing words weighted within the interval of $[-1; 1]$ plus their part of speech tag, and if applicable, their inflections. The current version of SentiWS (v1.8b) contains 1,650 negative and 1,818 positive words, which sum up to 16,406 positive and 16,328 negative word forms, respectively. It not only contains adjectives and adverbs explicitly expressing a sentiment, but also nouns and verbs implicitly containing one. The present work describes the resource's structure, the three sources utilised to assemble it and the semi-supervised method incorporated to weight the strength of its entries. Furthermore the resource's contents are extensively evaluated using a German-language evaluation set we constructed. The evaluation set is verified being reliable and its shown that SentiWS provides a beneficial lexical resource for German-language sentiment analysis related tasks to build on.

1. Introduction

An affect lexicon is a compendium of lexical entries for *affect* words with their corresponding parts of speech, affect categories, centralities, and intensities. An *affect word* is any word having an affect-related meaning or connotation. Any given affect word may have multiple entries in an affect lexicon, differing by its part of speech and/or its category (Subasic and Huettner, 2001). In these lexicons entries are labeled with their *prior polarity* (Wilson et al., 2009), i.e. their polarity without any given context or discourse. An affect lexicon may then be used to compute the frequency of sentiment bearing words in a text. Thereby the attempt usually is to percieve sentences comprising the unambiguous sense of the sentiment bearing word and/or taking into account *contextual valence shifters* (Polanyi and Zaenen, 2006).

1.1. Motivation and Related Work

In (Esuli and Sebastiani, 2006)'s *SentiWordNet* the authors assign positivity, negativity and objectivity values to WordNet-synsets. (Argamon et al., 2007)'s *Appraisal Lexicon* provides a source for appraisal adjectives, adverbs and adverb modifiers tagged with their attitude type and their semantic orientation. But just like SentiWordNet and the Appraisal Lexicon most of the resources in the domain of sentiment analysis publicly available for research are mainly Anglo-centric and therefore we believe there is a need for resources for languages other than English.

1.2. Outline

We first present the dictionary structure of SentiWS – a publicly available German-language resource for sentiment analysis[1] and the sources utilised to assemble it. We furthermore introduce the way we calculate the weight of an entry as an expression of its prior polarity with a value between -1.0 and $+1.0$. Finally we evaluate its performance, discuss the results and draw conclusions for further research.

2. Dictionary Structure

Entries in the dictionary schematically look like shown in Table 1.

Word	POS Tag	Weight	Inflections
harmonisch	ADJX[2]	+0.5243	harmonische, ..., harmonischst
Krise	NN	−0.3631	Krisen

Table 1: The Schema of SentiWS entries

The part of speech tags (POS tags) are given in the form of (Thielen et al., 1999)'s *Stuttgart-Tübingen-Tagset* (STTS). As POS tags are only provided for the baseforms, they are limited to those of adjectives, adverbs, normal nouns and infinite verbs. The inflections were, where available, retrieved from an internal database and are not guaranteed to be complete and error-free. Table 2 provides a comprehensive overview of the dictionary's content.

		Positive	Negative
Adjectives	Baseforms	784	698
	Inflections	11,782	10,604
Adverbs	Baseforms	6	4
	Inflections	0[3]	0[3]
Nouns	Baseforms	584	686
	Inflections	521	806
Verbs	Baseforms	312	430
	Inflections	2,453	3,100
All	Baseforms	1,650	1,818
	Inflections	14,756	14,510
	Total	**16,406**	**16,328**

Table 2: Overview of the dictionary's content

[1] http://asv.informatik.uni-leipzig.de/download

[2] This tag subsumes attributive and descriptive adjectives.

[3] German-language adverbs do not inflect.

3. Sources

SentiWS exploits several sources providing words plus their *semantic orientation* in one way or the other. All sources required a manual revision as described below.

3.1. General Inquirer

The first source is (Stone et al., 1966)'s *General Inquirer* (GI) lexicon. GI's categories *Pos* and *Neg* were semi-automatically translated into German using *Google Translate*[4] and manually revised afterwards, i.e. words were removed when inappropriate or without prior polarity. The reasons the authors chose the GI lexicon as a basis are its wide acceptance and that its coverage is comparably broad.
Apart from the words translated from the GI lexicon, a few hand-selected words from the domain of finance, e.g. *Finanzkrise* (i.e. *financial crisis*) and *Bankrott* (i.e. *insolvency*), were added to the basis of SentiWS, as it was originally developed for a study on the effects of financial newspaper articles and respective blog posts on a German stock index, the DAX 30, and vice versa (Remus et al., 2009).

3.2. Co-occurrence Analysis

The second source results from a special kind of *co-occurrence analysis* of rated product reviews provided by a business partner. Each review was tagged by its author to be either strongly positive or strongly negative. We added a positive or negative marker as an additional *pseudo-word* to each review and identified words which appear significantly often with one of these markers. 5,100 positively marked and 5,100 negatively marked reviews (containing 30,074 and 36,743 sentences, respectively) were used to carry out this co-occurrence analysis incorporating the *log-likelihood-measure* proposed by (Dunning, 1993), resulting in lists of word forms which significantly often appear together with one of the markers. These *candidate word forms* with positive or negative sentiment we manually inspected and chose from. The 200 most significant word forms have a precision of about 32% for the positive marker and 49.5% for the negative marker, respectively.
This kind of co-occurrence analysis yields a valuable source for *domain-dependent terminology*, i.e. sentiment expressions mostly used in specific contexts, e.g. product reviews. Significant co-occurrences we identified are for example *Reklamation* (i.e. a *customer complaint*) and *Fehlkauf* (i.e. a *mispurchase*), both clear expressions of negative sentiment.

3.3. German Collocation Dictionary

The third source is the forthcoming *German Collocation Dictionary* (Quasthoff, 2010). Among other things, this dictionary groups words that collocate with certain nouns by their *semantic similarity*. We used the words supplied by the two sources described above to distinguish between semantic groups *related to sentiment* and semantic groups not related to sentiment and thus were able to infer additional sentiment bearing words. At its current stage the German Collocation Dictionary contains 25,288 semantic groups with about 27.4% (6,932 groups) related to sentiment and about 0.003% (76 groups) strongly related to sentiment.
These groups are far from being disjoint, but provide some medium and low frequent words, e.g. *sonnendurchflutet* (i.e. *flooded by sunlight*), *umjubelt* (i.e. *highly acclaimed*), *glasklar* (i.e. *crystal-clear*) and *bärenstark* (i.e. *husky*), all expressions of positive sentiment.

[4] `http://translate.google.com`

4. Polarity Weighting

The weights mentioned above were retrieved utilising a method first suggested by (Church and Hanks, 1990): the so called *Pointwise Mutual Information* (PMI). This approach was successfully re-used for work related to sentiment analysis – the determination of the semantic orientation and its strength of adjectives – by (Turney, 2002) and (Turney and Littman, 2003). Their general strategy is to infer semantic orientation from *semantic association*. The semantic orientation SO of a given word w is calculated from the strength of its association A with a manually-selected set of positive seed words P minus the strength of its association with a set of negative seed words N (cf. Equation 1).

$$\text{SO-A}(w) = \sum_{p \in P} A(w,p) - \sum_{n \in N} A(w,n) \qquad (1)$$

The word w is classified as having a positive semantic orientation when $\text{SO-A}(w)$ is positive and a negative semantic orientation when $\text{SO-A}(w)$ is negative. The absolute value of $\text{SO-A}(w)$ can be considered the *strength* of its semantic orientation.
Parallel to (Turney and Littman, 2003)'s *paradigms* we used the following German seed sets P_{de} and N_{de}:

$$P_{de} = \left\{ \begin{array}{l} \text{gut, schön, richtig,} \\ \text{glücklich, erstklassig,} \\ \text{positiv, großartig, ausgezeichnet,} \\ \text{lieb, exzellent, phantastisch} \end{array} \right\} \qquad (2)$$

$$N_{de} = \left\{ \begin{array}{l} \text{schlecht, unschön, falsch,} \\ \text{unglücklich, zweitklassig,} \\ \text{negativ, scheiße, minderwertig,} \\ \text{böse, armselig, mies} \end{array} \right\} \qquad (3)$$

The semantic associations $A(w,p)$ and $A(w,n)$ are then calculated using the PMI. The PMI between two words w_1 and w_2 according to (Church and Hanks, 1990) is defined as given in Equation 4,

$$\text{PMI}(w_1, w_2) = \log_2 \left(\frac{P(w_1 \,\&\, w_2)}{P(w_1) \cdot P(w_2)} \right) \qquad (4)$$

where $P(w)$ is the probability that w occurs and $P(w_1 \& w_2)$ is the probability that w_1 and w_2 co-occur. These probabilities were estimated using frequencies and co-occurrence statistics on a internal German-language corpus consisting of approximately 100 Million sentences.
If $\text{SO-A}(w)$ of a word that was entered as being positive is negative or vice versa, the word in question is either removed, put in the opposite class, or, if after manually revision we find its classification is correct, its weight is set to the minimum weight of its class. All weights are scaled to

the interval of $[-1; 1]$ and rounded to 4 decimal places with $+1.0$ being absolutely positive and -1.0 being absolutely negative.
Very positive words are for example *Freude* (i.e. *joy*) with a weight of 0.6502 and *perfekt* (i.e. *perfect*) with a weight of 0.7299. Very negative words are for example *betrügen* (i.e. to *betray*) with a weight of -0.743 and *schädlich* (i.e. *harmful*) with a weight of -0.9269.
The distribution of the absolute weights in SentiWS (cf. Figure 1) follows a Zipf-like distribution (Zipf, 1972): Very little word forms have high weights, some word forms have medium weights and a large amount of word forms have little or very little weights.

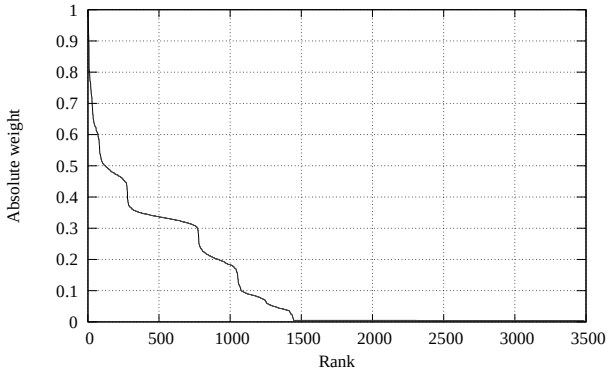

Figure 1: The distribution of the absolute weights

5. Evaluation

Just like there was no German-language dictionary for sentiment analysis, there is no corresponding data set for evaluation purposes. In order to evaluate SentiWS' performance we compiled such a data set. We first randomly selected 2,000 sentences from a corpus containing posts from a variety of internet fora and then manually categorized them as being positive, negative or neutral. We then randomly selected 160 sentences from each category, resulting in a data set of 480 sentences. The minimum sentence length is 4 word forms, the maximum sentence length is 40 word forms and the approximate average sentence length is 15.97 word forms.
Two human raters, one of which is an author of this paper, were then instructed to *annotate* each sentence regarding the prior polarities of each adjective, adverb, noun or verb in it, i.e. they had to decide whether a word form was positive, negative or had no prior polarity. Measuring the raters' overall agreement using Cohens κ (Cohen, 1960) in a free-marginal variant (Brennan and Prediger, 1981) the interrater reliability is $\kappa_{free} = 0.76$ and thus is considered being reliable.
Hereupon the sentences were preprocessed incorporating the *Stanford POS Tagger 2.0*[5] and, taking into account the POS tags, the raters' annotations were compared with the entries in SentiWS. Errors induced by the Stanford POS Tagger were excluded and precision P, recall R and f-measure F were calculated as shown in Table 3.
SentiWS' entries were checked against the annotations of rater 1, rater 2 and their *consensus*, i.e. the annotations in which both raters agreed. All results are given for positive word forms only, negative word forms only and all word forms. Generally SentiWS performs better identifying negative word forms ($F = 0.86$) than it does for positive word forms ($F = 0.82$), but the overall performance is very promising ($P = 0.96$, $R = 0.74$, $F = 0.84$).
Typical errors that lower the recall are missing words and missing word forms, for example domain-specific terms (e.g. *Kantigkeit*, i.e. *"edgy-ness"*), foreign words, colloquial language (e.g. *Ka****, i.e. *sh***) and mistypes or misspellings. Typical errors that lower the precision include words that are ambiguous and tend to be polar in one sense but not the other.

6. Further Work

SentiWS is *work in progress* and hence far from being fully-fledged and error-free. It will be continuously refined by adding missing words and word forms and removing ambiguous ones. It is furthermore likely that it will be extended by introducing a new dimension indicating *subjectivity*, just like (Esuli and Sebastiani, 2006)'s SentiWordNet does. Apart from that the authors recently took interest in representing more fine-grained emotions, aside from pure polarity (Whitelaw et al., 2005).
We also believe it is necessary to delve into weighting schemes. Although we used the PMI without questioning it, we are very aware of the fact that the weighting itself needs to be evaluated and possibly contrasted with other weighting methods (Landauer and Dumais, 1997; Richardson et al., 1994; Budanitsky and Hirst, 2001; Biemann, 2006)

7. Summary

We have in detail presented a German-language affect dictionary which attributes each word with its syntactic category, its inflectional forms, its polarity and its strength. We conducted an evaluation and proved SentiWS being a useful resource for sentiment analysis related tasks to build on. As far as we know SentiWS is the first German-language dictionary dedicated to sentiment analysis, opinion mining etc. publicly available and we encourage researchers to use it in composition with the other corpora, tools and webservices provided by the *Wortschatz* project[6] (Quasthoff et al., 2006; Biemann et al., 2007; Biemann et al., 2008; Büchler and Heyer, 2009)

8. References

S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani. 2007. Automatically Determining Attitude Type and Force for Sentiment Analysis. In *Proceedings of the 3rd Language & Technology Conference (LTC-07)*.

C. Biemann, G. Heyer, U. Quasthoff, and M. Richter. 2007. The Leipzig Corpora Collection – Monolingual Corpora of Standard Size. In *Proceedings of the 4th Conference on Corpus Linguistics (CL2007)*.

[5] http://nlp.stanford.edu/software/tagger.shtml [6] http://wortschatz.informatik.uni-leipzig.de

	Pos			Neg			All		
	P	R	F	P	R	F	P	R	F
Rater 1	0.94	0.70	0.80	0.99	0.74	0.85	0.96	0.72	0.82
Rater 2	0.89	0.62	0.73	0.97	0.72	0.83	0.92	0.66	0.77
Consensus	0.94	0.72	0.82	0.99	0.76	0.86	**0.96**	**0.74**	**0.84**

Table 3: Evaluation results given as precision P, recall R and f-measure F

C. Biemann, U. Quasthoff, G. Heyer, and F. Holz. 2008. ASV Toolbox – A Modular Collection of Language Exploration Tools. In *Proceedings of the 6th International Language Resources and Evaluation (LREC'08)*, pages 1760–1767.

C. Biemann. 2006. Unsupervised Part-of-Speech Tagging Employing Efficient Graph Clustering. In *Proceedings of the 21st International Conference on Computational Linguistics (COLING) / Association for Computational Linguistics (ACL) Student Research Workshop*, pages 7–12.

R.L. Brennan and D.J. Prediger. 1981. Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement*, 41(3):687–699.

A. Budanitsky and G. Hirst. 2001. Semantic Distance in WordNet: an Experimental, Application-oriented Evaluation of five Measures. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL) Workshop on WordNet and Other Lexical Resources*, pages 29–34.

M. Büchler and G. Heyer. 2009. Leipzig Linguistic Services – A 4 Years Summary of Providing Linguistic Web Services. In G. Heyer, editor, *Proceedings of the Conference on Text Mining Services (TMS)*.

K.W. Church and P. Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, 16(1):22–29.

J. Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.

T. Dunning. 1993. Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1):61–74.

A. Esuli and F. Sebastiani. 2006. SentiWordNet: A publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th International Language Resources and Evaluation (LREC'06)*, pages 417–422.

T.K. Landauer and S.T. Dumais. 1997. A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 104(2):211–240.

L. Polanyi and A. Zaenen. 2006. Contextual Valence Shifters. In J. Shanahan, Y. Qu, and J. Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Application*, volume 20 of *The Information Retrieval Series*, pages 1–9.

U. Quasthoff, M. Richter, and C. Biemann. 2006. Corpus Portal for Search in Monolingual Corpora. In *Proceedings of the 5th International Language Resources and Evaluation (LREC'06)*, pages 1799–1802.

U. Quasthoff. 2010. *Deutsches Kollokationswörterbuch*. deGruyter, Berlin, New York.

R. Remus, K. Ahmad, and G. Heyer. 2009. Sentiment in German-language News and Blogs, and the DAX. In G. Heyer, editor, *Proceedings of the Conference on Text Mining Services (TMS)*, pages 149–158.

R. Richardson, A.F. Smeaton, and J. Murphy. 1994. Using WordNet as a Knowledge Base for Measuring Semantic Similarity Between Words. In *Proceedings of the 7th Irish Conference on Artificial Intelligence and Cognitive Science (AICS)*.

P.J. Stone, D.C. Dunphy, M.S. Smith, D.M. Ogilvie, et al. 1966. *The General Inquirer: a Computer Approach to Content Analysis*. MIT Press.

P. Subasic and A. Huettner. 2001. Affect Analysis of Text using Fuzzy Semantic Typing. *IEEE Transactions on Fuzzy Systems*, 9(4):483–496.

C. Thielen, A. Schiller, S. Teufel, and C. Stöckert. 1999. Guidelines für das Tagging deutscher Textkorpora mit STTS. Technical report, University of Stuttgart and University of Tübingen.

P.D. Turney and M.L. Littman. 2003. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *Association for Computing Machinery (ACM) Transactions on Information Systems (TOIS)*, 21(4):315–346.

P.D. Turney. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424.

C. Whitelaw, N. Garg, and S. Argamon. 2005. Using Appraisal Groups for Sentiment Analysis. In *Proceedings of the 14th Association for Computing Machinery (ACM) International Conference on Information and Knowledge Management (ICIKM)*, pages 625–631.

T. Wilson, J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: an Exploration of Features for Phrase-level Sentiment Analysis. *Computational Linguistics*, 35(3):399–433.

G.K. Zipf. 1972. *Human Behavior and the Principle of Least Effort*. Hafner, New York.