GermanPolarityClues: A Lexical Resource for German Sentiment Analysis.

Conference Paper · January 2010

Source: DBLP

1 author:

Ulli Waltinger
Siemens

GermanPolarityClues: A Lexical Resource for German Sentiment Analysis

Ulli Waltinger

Text Technology, Bielefeld University, Universitätsstrasse 3, 33602 Bielefeld, Germany
ulli_marc.waltinger@uni-bielefeld.de

Abstract

In this paper, we propose *GermanPolarityClues*, a new publicly available lexical resource for sentiment analysis for the German language. While sentiment analysis and polarity classification has been extensively studied at different document levels (e.g. sentences and phrases), only a few approaches explored the effect of a polarity-based feature selection and subjectivity resources for the German language. This paper evaluates four different English and three different German sentiment resources in a comparative manner by combining a polarity-based feature selection with SVM-based machine learning classifier. Using a semi-automatic translation approach, we were able to construct three different resources for a German sentiment analysis. The manually finalized *GermanPolarityClues* dictionary offers thereby a number of $10,141$ polarity features, associated to three numerical polarity scores, determining the positive, negative and neutral direction of specific term features. While the results show that the size of dictionaries clearly correlate to polarity-based feature coverage, this property does not correlate to classification accuracy. Using a polarity-based feature selection, considering a minimum amount of prior polarity features, in combination with SVM-based machine learning methods exhibits for both languages the best performance (F1: 0.83-0.88).

1. Introduction

Sentiment analysis refers to a discipline of information retrieval - the opinion mining (OM). OM analyzes the characteristics of opinions, feelings and emotions that are expressed in textual (Pang et al., 2002; Dave et al., 2003; Hu and Liu, 2004; Wilson et al., 2005; Annett and Kondrak, 2008) or spoken (Becker-Asano and Wachsmuth, 2009) data with respect to a certain subject. A subtask of sentiment analysis, which has been extensively studied in recent years, is the sentiment categorization on the basis of certain polarities - the sentiment polarity identification (Pang et al., 2002). This task focuses on the classification of positive, negative or neutral expressions in texts. With respect to the task of polarity-related term feature interpretation, most of the proposed methods make use of manually annotated or automatically constructed lists of subjectivity terms. While there are a various resources and data sets proposed in the research community, only a small number are freely available to the public – most of them for the English language. For the German language, there is, to the best of our knowledge, currently no annotated dictionary (terms with their associated semantic orientation) freely available.

In this paper, we propose *GermanPolarityClues*, a new publicly available lexical resource for sentiment analysis for the German language. We empirically show that a German-based feature selection on the basis of the newly created resource contributes to an automatic sentiment analysis.

2. Related Work

In recent years, various approaches have been proposed to the domain of sentiment analysis, either focusing on polarity-based feature selection methods (Tan and Zhang, 2008), such as document frequency, chi square or polarity selection, or combining rule-based (Pang et al., 2002; Turney and Littman, 2002; Kennedy and Inkpen, 2006), supervised and unsupervised classification methods (Chaovalit and Zhou, 2005; Prabowo and Thelwall, 2009), such as $k-NN$, Naive Bayes and Support Vector Machine (SVM). In general, most of the published evaluations indicate that the combination of a sentiment-based feature selection (using polarity features only) and machine learning algorithms on the basis of SVM produces the best performance with respect to classification accuracy. However, at the center of nearly all approaches, an external resource is used in order to detect and extract polarity-related term features in text.

With respect to the used sentiment or subjectivity resources, only a few of them are publicly available, mostly inducing the English language. (Hatzivassiloglou and McKeown, 1997) used a small set of manually annotated ($1,336$ adjectives) in order to extract polarity-related adjectives using a bootstrapping strategy, inducing *adjective conjunction* ($13,426$) that hold the same semantic orientation. Various resources used the linguistic resource *WordNet* (Fellbaum, 1998) as the basis for the construction of sentiment resources, inducing graph-related distance measures (Maarten et al., 2004), classifying word-to-synset relations (Strapparava and Valitutti, 2004) (*WordNet-Affect* comprises $2,874$ synsets and $4,787$ words) or combining semantic relations with co-occurrence information extracted from corpus using the *Ising Spin Model* (Chandler, 1987, pp. 119) (*SentiSpin* induces $88,015$ words) (Takamura et al., 2005). Also on the basis of *WordNet*, (Esuli and Sebastiani, 2006) proposed a method for the analysis of glosses and associated synset (*SentiWordNet* comprises $144,308$ terms). (Wiebe et al., 2005; Wilson et al., 2005; Wiebe and Riloff, 2005) presented the most fine-grained polarity resource. In total, 8,221 term features were not only rated by their polarity (positive, negative, both, neutral) but also by their reliability (e.g. strongly subjective, weakly subjective). Most recently, (Waltinger, 2009) proposed an approach of term-based polarity enhancement comprising social network properties (Mehler, 2008). Using the entries of the *SpinModel* dataset as seed words, associated phrase and term definitions were extracted from the *urban dictionary* project (*Polarity Enhancement*: $137,088$ term).

3. Methodology

The method we have used to build *GermanPolarityClues* is a semi-automatic translation approach of existing English-based sentiment resources to the German language. Different to the approach of (Denecke, 2008), by translating a German input text into the English language (*SentiWordNet* as a resource), we rather focused on building a new German dictionary by translating polarity features only. Since existing resources vary significantly in the number of comprised polarity term features ($6,663 - 144,308$), we approached the construction of the new resource in three steps:
First, we systematically evaluated the most widely used English-based sentiment resources (Subjectivity Clues (Wiebe et al., 2005), SentiSpin (Takamura et al., 2005), SentiWordNet (Esuli and Sebastiani, 2006) and Polarity Enhancement (Waltinger, 2009)) in a document-based polarity identification experiment (Waltinger, 2010). That is, we analyzed how the different subjectivity resources perform within the same experimental setup. Does the significant difference in quantity of used polarity features affect the performance of opinion mining?
Second, we translated the two most comprehensive dictionaries, the *Subjectivity Clues* (Wiebe et al., 2005; Wilson et al., 2005; Wiebe and Riloff, 2005) comprising $9,827$ term features (further called *German Subjectivity Clues*) and the *SentiSpin* (Takamura et al., 2005) dictionary, comprising $105,561$ polarity features (further called *German SentiSpin*), into the German language by automatic means. More precisely, we have translated each English polarity feature into the German language using an English-to-German translation software[1]. While there are in many cases more than one possible translations available, we decided to take a maximum number of three translations for dictionary construction into account. Therefore, the size of the built German resources differ to their English pendant. With respect to polarity feature weights, each aggregated German feature has inherited the sentiment orientation score (e.g. positive, negative, neutral) of the initial seed word from the English resource (e.g. English: "brave"—"positive" $\mapsto$ German: "mutig"—"positive"). This approach clearly leads to a problem of term ambiguity. We therefore decided to compile in a third step the *GermanPolarityClues* dictionary, by *manually* assessing each individual term feature of the *German Subjectivity Clues* dataset by their sentiment orientation (See Table 2.). In addition, we added to this resource a number of 290 German negation-phrases (e.g. "nicht schlecht" = "not bad") and the most frequent positive and negative synonyms of existing term features, which previously had not been in there[2] - inducing a total size of $10,141$ polarity features (see Table 3.). Finally, we conducted an extensive evaluation on the three constructed German resources in a comparative manner by means of a SVM-based polarity classification setup.

[1]We have used the online service of dict.leo.org for the translation of term features.

[2]Note, these features were extracted from the dataset of the de.wiktionary.org project.

Overall Features:	10,141
No. Positive Features:	3,220
No. Negative Features:	5,848
No. Neutral Features:	1,073
No. Negation Features:	290
No. Noun Features:	4,408
No. Verb Features:	2,728
No. Adj/Adv Features:	2,604

Table 5: GermanPolarityClues feature statistics by polarity and grammatical categories.

4. Experiments

As stated above, since there are no published baseline results for a sentiment polarity identification experiment for the German language, we chose to analyze the quality of the English-based polarity resources as reference line. That is, we first used each of the respective sentiment resources (German and English) for a polarity-related feature selection. Second, we applied a document-based hard-partition machine learning classifier (Pang et al., 2002; Chaovalit and Zhou, 2005; Tan and Zhang, 2008; Prabowo and Thelwall, 2009; Waltinger, 2009) using Support Vector Machines (SVM) (Joachims, 2002b) (SVM^{Light} V6.01 (Joachims, 2002a)) for the task of sentiment polarity classification. In each case of the SVM-Classifiers, *Linear*- and *RBF-Kernel* were evaluated in a comparative manner.
Since our experiments comprise two different languages, we have used two different evaluation corpora. For the English language we conducted the polarity identification classification (Waltinger, 2010) using the movie review corpus, initially compiled by (Pang et al., 2002). This corpus consists of two polarity categories (positive and negative), each category comprises 1000 articles with an average of 707.64 textual features. With respect to the German language, we manually created a reference corpus by extracting review data from the *Amazon.com* website (see Figure 1). Contributed reviews at *Amazon.com* correspond to human-rated product reviews with an attached rating scale from 1 (worst) to 5 (best) stars. For the experiment, we have used 1000 reviews for each of the 5 ratings, each comprising 5 different categories. All category, star label and authorship information were removed from the documents. The average number of term features of the comprised reviews was 109.75. With respect to the experiments on the German corpus, we evaluated different "Star" combinations as positive and negative categories (e.g classifying Star1 against Star5, but also Star1 and Star2 against Star 4 and Star 5). Note, we conducted the experiments using a sentiment-based feature selection only. That is, we did not evaluate the quality of polarity orientation, but rather used the constructed dictionaries as a resource for a polarity feature selection. Subsequently, all identified features were weighted by the $tf - idf$ schema (Salton and McGill, 1983). We report the $F1$-Measure as calculated by the $leave - one - out$ cross-validation of SVM^{light}. With respect to the polarity orientation scores of the built sentiment dictionaries, we additionally used the Amazon-Corpus in order to obtain corpus-based polarity scores (see

Id:	Feature	PoS	A(+)	A(−)	A(∘)	B(+)	B(−)	B(∘)
5653	Begründung	NN	0	0	1	0	0.5	0.5
7573	Katastrophe	NN	0	1	0	0	0.68	0.32
7074	ideal	ADJD	1	0	0	0.76	0.13	0.11

Table 1: Overview of the GermanPolarityClues data schema by (A) automatic- and (B) corpus-based polarity orientation rating.

Rank	N-Feature	N-Frequency	V-Feature	V-Frequency	A-Feature	A-Frequency
1	Ding	174	müssen	1186	nur	2146
2	Fehler	112	enttäuschen	283	klein	708
3	Nachteil	97	fallen	193	leid	537
4	Abzug	66	fehlen	137	schlecht	448
5	Druck	60	brechen	83	alt	401
6	Enttäuschung	60	aufgeben	63	leider	342
7	Gewicht	59	bereuen	54	kurz	339
8	Mangel	56	verlassen	34	fast	265
9	Gegensatz	45	ärgern	34	teuer	210
10	Versuch	44	abbrechen	30	kaum	161

Table 2: Negative polarity features by corpus rank, frequency and grammatical category (Noun, Verb, Adjective/Adverb)

Figure 1: Overview of the product review section at Amazon.com using a ”Star”-based rating scale.

Table 2.). Thereby, we calculated the polarity probability of each term feature by means of their human-created online rating, using ”Star1-2” reviews as a negative, ”Star 3” as a neutral, and ”Star 4-5” reviews as a positive polarity indication (e.g. occurrences of term *hoffnungsvoll* in sub-corpus ”Star1-2” divided by the number of occurrences within the entire corpus) .

5. Results

The results (Table 5.) for the English-based baseline experiments (see the preliminary study of (Waltinger, 2010)) indicate, that the smallest resource, *Subjectivity Clues*, perform with a touch better than *SentiWordNet*, *SentiSpin* and the *Polarity Enhancement*. dataset (F1-Measure results range between $82.9 - 83.9$). At this stage, we can argue that a subjectivity feature selection in combination with machine learning classifier clearly outperform the well known baseline results as published by (Pang et al., 2002) (Naive Bayes: $acc = 78.7$; Maximum Entropy: $acc = 81.0$; N-Gram-based SVM: $acc = 82.9$). Interestingly, even the biggest dictionary with the highest coverage property does not outperform the resource with the lowest number of polarity-features. Starting from these preliminary findings, we are using the English results as a reference line for the assessment of the German sentiment resources. Overall, the results of the newly build German subjectivity resources (see Table 3.), used for the document-based polarity identification, indicate similar perceptions. Using the *German SentiSpin* version, comprising $105,561$ polarity features, lets us gain a promising F1-Measure of 85.9. The *German Subjectivity Clues* dictionary, comprising $9,827$ polarity features, performs with an F1-Measure of 84.1 almost at the same level. However, the *GermanPolarityClues* dictionary, comprising $10,141$ polarity features, outperforms with an F1-Measure of 87.6 all other German resources. In addition, with respect to the number of polarity features actually used within the Amazon-based SVM-classification experiments (see Table 5.), we can identify that a number of $2,700$ features only within the *GermanPolarityClues* dictionary exhibits the best performance. It seems that this newly created sentiment resource, which induces a rather small feature size (10-times smaller than the *German SentiSpin*), is due to its manual controlled vocabulary and its introduced negation- and synonym-pattern, of high-quality for the task of polarity identification. Thus, we argue that the newly created and freely available[3] *GermanPolarityClues* dictionary is a promising resource for a German-based sentiment analysis.

German SentiSpin:	10,802
German Subjectivity:	2,657
German Polarity Clues:	2,700

Table 8: Number of polarity features used for the SVM-Classification by comprised resources.

[3]The constructed resources can be freely accessed and downloaded at: http://hudesktop.hucompute.org/

Rank	N-Feature	N-Frequency	V-Feature	V-Frequency	A-Feature	A-Frequency
1	Super	565	geraten	174	gut	3354
2	Leistung	223	klingen	163	sehr	2172
3	Einsatz	131	begeistern	131	mehr	1189
4	Spa	126	erhalten	115	einfach	1071
5	Ergebnis	122	wunderbar	75	viel	1014
6	Dank	75	überraschen	62	ganz	718
7	Freude	61	verdienen	57	neu	701
8	Empfehlung	58	ankommen	50	schnell	670
9	Wert	57	bestehen	46	groß	605
10	Gefül	56	genießen	39	lang	567

Table 3: Positive polarity features by corpus rank, frequency and grammatical category (Noun, Verb, Adjective/Adverb)

Resource:	Subject. Clues	Senti Spin	Senti WordNet	Polarity Enhance	German SentiSpin	German Subject.	German Polarity Clues
No. of Features:	6,663	88,015	144,308	137,088	105,561	9,827	10,141
Positive-AMean:	76.83	236.94	241.36	239.25	53.63	27.70	26.66
Positive-StdDevi:	30.81	84.29	85.61	84.98	6.90	4.59	5.01
Negative-AMean:	69.72	218.46	223.11	221.25	50.18	25.68	24.14
Negative-StdDevi:	26.22	74.08	75.37	74.68	10.40	5.88	5.41
Text-AMean:	707.64	707.64	707.64	707.64	109.75	109.75	109.75
Text-StdDevi:	296.94	296.94	296.94	296.94	24.52	24.52	24.52

Table 4: The standard deviation (StdDevi) and arithmetic mean (AMean) of subjectivity features by resource, text corpus (Text) and polarity category (Positive, Negative) (Waltinger, 2010).

6. Conclusions

In this paper, we proposed a new publicly available lexical resource for sentiment analysis for the German language - *GermanPolarityClues*. The new resource was built combining a semi-automatic translation method and a manually assessment and extension of individual polarity-based term features. We empirically showed that the *GermanPolarityClues* dictionary can be, with an F1-Measure of 87.6, a valuable resource for a polarity-based feature selection. However, the current study can only be seen as a starting point in the construction of resources for a German-based sentiment analysis. Future work includes the extension and revalidation of the existing dataset with additional polarity features as aggregated from other (web-based) resources and dictionaries. We also plan to conduct an human-judgement-based assessment of the other two resources, in order to improve the existing *GermanPolarityClues* dictionary.

7. Acknowledgment

We gratefully acknowledge financial support of the German Research Foundation (DFG) through the EC 277 *Cognitive Interaction Technology* at Bielefeld University.

8. References

Michelle Annett and Grzegorz Kondrak. 2008. A comparison of sentiment analysis techniques: Polarizing movie blogs. In *Canadian Conference on AI*, pages 25–35.

Christian Becker-Asano and Ipke Wachsmuth. 2009. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*.

David Chandler. 1987. *Introduction to Modern Statistical Mechanics*. Oxford University Press.

Pimwadee Chaovalit and Lina Zhou. 2005. Movie review mining: a comparison between supervised and unsupervised classification approaches. *Hawaii ICSS*, 4.

Kushal Dave, Steve Lawrence, and David M. Pennock. 2003. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proc. of WWW '03*, pages 519–528. ACM Press.

Kerstin Denecke. 2008. Using sentiwordnet for multilingual sentiment analysis. In *ICDE Workshops*, pages 507–512. IEEE Computer Society.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proc. of the LREC06*, pages 417–422.

Christiane Fellbaum, editor. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proc. of the eighth EACL*, pages 174–181. ACL.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proc. of the tenth ACM SIGKDD KDD '04*, pages 168–177, New York, NY, USA. ACM.

T. Joachims. 2002a. SVM light, http://svmlight.joachims.org.

Thorsten Joachims. 2002b. *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Kluwer Academic Publishers.

Alistair Kennedy and Diana Inkpen. 2006. Sentiment clas-

Resource	Model	F1-Positive	F1-Negative	F1-Average
German SentiSpin Star1+2 vs. Star4+5	SVM-Linear	.827	.828	.828
	SVM-RBF	.830	.830	**.830**
German SentiSpin Star1 vs. Star5	SVM-Linear	.857	.861	**.859**
	SVM-RBF	.855	.858	.857
German Subjectivity Star1+2 vs. Star4+5	SVM-Linear	.810	.813	**.811**
	SVM-RBF	.804	.803	.803
German Subjectivity Star1 vs. Star5	SVM-Linear	.841	.842	**.841**
	SVM-RBF	.834	.834	.834
GermanPolarityClues Star1+2 vs. Star4+5	SVM-Linear	.875	.730	**.803**
	SVM-RBF	.866	.661	.758
GermanPolarityClues Star1 vs. Star5	SVM-Linear	.875	.876	**.876**
	SVM-RBF	.855	.850	.853

Table 6: F1-Measure evaluation results of a German subjectivity feature selection using SVM.

Resource	Model	F1-Positive	F1-Negative	F1-Average
English Subjectivity Clues	SVM-Linear	.832	.823	**.828**
	SVM-RBF	.828	.823	.826
English SentiWordNet	SVM-Linear	.832	.828	**.830**
	SVM-RBF	.816	.812	.814
English SentiSpin	SVM-Linear	.831	.827	**.829**
	SVM-RBF	.815	.811	.813
English Polarity Enhancement	SVM-Linear	.841	.837	**.839**

Table 7: F1-Measure evaluation results of an English subjectivity feature selection using SVM. (Waltinger, 2010)

sification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125.

Jaap Kamps Maarten, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. 2004. Using wordnet to measure semantic orientations of adjectives. In *National Institute for*, pages 1115–1118.

Alexander Mehler. 2008. A short note on social-semiotic networks from the point of view of quantitative semantics. In Harith Alani, Steffen Staab, and Gerd Stumme, editors, *Proc. of the Dagstuhl Seminar on Social Web Communities, September 21-26, Dagstuhl*.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proc. of the ACL-02 EMNLP '02*, pages 79–86. ACL.

Rudy Prabowo and Mike Thelwall. 2009. Sentiment analysis: A combined approach. *J. Informetrics*, 3(2):143–157.

G. Salton and M. McGill. 1983. *Introduction to Modern Information Retrieval*. McGraw-Hill, New York.

C. Strapparava and A. Valitutti. 2004. WordNet-Affect: an affective extension of WordNet. In *Proc. of LREC*, volume 4, pages 1083–1086.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proc. of ACL '05*, pages 133–140. ACL.

Songbo Tan and Jin Zhang. 2008. An empirical study of sentiment analysis for chinese documents. *Expert Syst. Appl.*, 34(4):2622–2629.

Peter D. Turney and Michael L. Littman. 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *CoRR*, cs.LG/0212012.

Ulli Waltinger. 2009. Polarity reinforcement: Sentiment polarity identification by means of social semantics. In *Proc. of the IEEE Africon 2009, September 23-25, Nairobi, Kenya*.

Ulli Waltinger. 2010. Sentiment analysis reloaded: A comparative study on sentiment polarity identification combining machine learning and subjectivity features. In *Proceedings of the 6th International Conference on Web Information Systems and Technologies (WEBIST '10)*, Valencia, Spain, April.

Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. In *Proc. of CICLing-05*, volume 3406, pages 475–486, Mexico City, MX. Springer-Verlag.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 1(2):0.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. of the HLT '05*, pages 347–354. ACL.