SENTIMENT ANALYSIS RELOADED
A Comparative Study on Sentiment Polarity Identification Combining Machine Learning and Subjectivity Features

Ulli Waltinger
Text Technology, Bielefeld University, Universitatsstrasse 3, 33602 Bielefeld, Germany

Keywords: Machine learning, Support vector machine, Sentiment analysis, Polarity identification, Subjectivity resources.

Abstract: This paper presents an empirical study on machine learning-based sentiment analysis. Though polarity classification has been extensively studied at different document-structure levels (e.g. document, sentence, words), little work has been done investigating feature selection methods and subjectivity resources. We systematically analyze four different English subjectivity resources for the task of sentiment polarity identification. 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 polarity-based feature selection, considering a minimum amount of prior polarity features, in combination with SVM-based machine learning methods exhibits the best performance ($acc = 84.1, f1 = 83.9$), in comparison to the classical approaches on polarity identification. Based on the findings of the English-based experimental setup, a new German subjectivity resource is proposed for the task of German-based sentiment analysis. The results of the experiments show, with $f1 = 85.9$ its good adaptability to the new domain.

1 INTRODUCTION

With the enormous growth of digital content arising in the web, document classification and categorization receives more and more interest in the information retrieval community. This relates to content-based models (Joachims, 2002a) as well as to structure-orientated approaches (Mehler et al., 2007). While a majority of approaches focusses on a thematical or topical differentiation of textual data, the task of sentiment analysis (Pang and Lee, 2008) refers to the (non-topical) opinion mining. This area focuses on the detection and extraction of opinions, feelings and emotions in text with respect to a certain subject. A subtask of this area, which has been extensively studied, is the sentiment categorization on the basis of certain polarities. That is, being able to distinguish between positive, neutral or negative expressions or statements of extracted textual (Pang et al., 2002; Dave et al., 2003; Hu and Liu, 2004; Wilson et al., 2005; Annett and Kondrak, 2008) or spoken elements (Becker-Asano and Wachsmuth, 2009). Moreover, finer-grained methods additionally explore the level or intensity of polarity inducing a rating inference (e.g. a rating scale between one and five stars) model. In the majority of approaches on sentiment polarity identification, the determination of subjectivity or polarity-related term features is in the center in order to draw conclusions about the actual polarity-related orientation of the entire text. Since positive as well as negative expressions can occur within the same document, this task is challenging. Considering the following example of an Amazon product review:

> Product-Review[1]: <u>Wonderful</u> when it works... I owned this TV for a month. At first I thought it was <u>terrific</u>. <u>Beautiful</u> <u>clear</u> picture and <u>good</u> sound for such a <u>small</u> TV. Like others,however, I found that it did not always retain the programmed stations and then had to be reprogrammed every time you turned it off. I called the manufacturer and they <u>admitted</u> this is a <u>problem</u> with the TV.

Although most of the polarity-related text features contribute to a positive review (e.g. wonderful, terrific, beautiful...), this user-contribution is classified as a negative review. This example clearly shows that classical text categorization approaches (e.g. bag-of-words) need to be extended or seized to the domain

[1] http://www.amazon.com/

Waltinger U.
SENTIMENT ANALYSIS RELOADED - A Comparative Study on Sentiment Polarity Identification Combining Machine Learning and Subjectivity Features.
DOI: 10.5220/0002772602030210
In *Proceedings of the 6th International Conference on Web Information Systems and Technology (WEBIST 2010)*, page
ISBN: 978-989-674-025-2

of sentiment analysis. Though, we consider polarity identification as a binary classification task, the determination of semantically oriented linguistic features on different structural levels (words, sentences, documents,...) is at the core of attention. With respect to the task of term feature interpretation, most of the proposed unsupervised or (semi-)supervised sentiment-related approaches make use of annotated and constructed lists of subjectivity terms.

While there are 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. In terms of coverage rate, the number of comprised subjectivity terms of these dictionaries varies significantly - ranging between $8,000$ and $140,000$ features. For the German language, there is, to the best of our knowledge, currently no annotated dictionary (terms with their associated semantic orientation) freely available. The questions that arise therefore are: How does the significant coverage variations of the English sentiment resources correlate to the task of polarity identification? Are there notable differences in the accuracy performance, if those resources are used within the same experimental setup? How does sentiment term selection combined with machine learning methods affect the performance? And finally, are we able to draw conclusions from the results of the experiments in building a German sentiment analysis resource?

In this paper, we investigate the effect of sentiment-based feature selection combined with machine learning algorithms in a comparative experiment, comprising the four most widely used subjectivity dictionaries. We empirically show that a sentiment-sensitive feature selection contributes to the task of polarity identification. Further, we propose based on the findings a subjectivity dictionary for the German language, that will be freely available to the public.

2 RELATED WORK

In this section, we present related work on sentiment analysis. A focus is set on comparative studies and different algorithms applied to the task of polarity identification. Tan and Zhang (2008) presented an empirical study of sentiment categorization on the basis of different feature selection (e.g. document frequency, chi square, subjectivity terms) and different learning methods (e.g k-nearest neighbor, Naive Bayes, SVM) on a Chinese data set. The results indicated that the combination of sentimental feature selection and machine learning-based SVM performs best compared to other tested sentiment classifiers.

Chaovalit and Zhou (2005) published a comparative study on supervised and unsupervised classification methods in a polarity identification scenario of movie reviews. Their results confirmed also that machine learning on the basis of SVM are more accurate than any other unsupervised classification approaches. Hence, a significant amount of training and building associated models is needed.

Prabowo and Thelwall (2009) proposed a combined approach for sentiment analysis using rule-based, supervised and machine learning methods. An overview of current sentiment approaches is given, compared by their model, data source, evaluation methods and results. However, since most of the current attempts based their experiments on different setups, using mostly self-prepared corpora or subjectivity resources, a uniform comparison of the proposed algorithms is barely possible. The results of the combined approach show that no single classifier outperforms the other, and the hybrid classifier *can* result in a better effectiveness.

With respect to different methods applied to the sentiment polarity analysis, we can identify two different branches. On the one hand - rule-based approaches, as for instance counting positive and negative terms (Turney and Littman, 2002) on the basis of semantic lexicon, or combining it with so called discourse-based contextual valence shifters (Kennedy and Inkpen, 2006). On the other hand - machine-learning approaches (Turney, 2001) on different document levels, such as the entire documents (Pang et al., 2002), phrases (Wilson et al., 2005; Taboada et al., 2009; Agarwal et al., 2009), sentences (Pang and Lee, 2004) or on the level of words (Maarten et al., 2004), using extracted and enhanced linguistic features from internal (e.g. PoS- or text phrase information) and/or external resources (e.g. syntactic and semantic relationships extracted from lexical resources such as WordNet (Fellbaum, 1998)) (Mullen and Collier, 2004; Chaovalit and Zhou, 2005). Most notably, sentence-based models have been quite intensively studied in the past, combining machine learning and unsupervised approaches using inter-sentence information (Yu and Hatzivassiloglou, 2003; Kugatsu Sadamitsu and Yamamoto, 2008), sentence-based linguistic feature enhancement (Wiegand and Klakow,) or most famous by following a sentence-based minimum cut strategy (Pang and Lee, 2004; Pang and Lee, 2005).

In general, sentence-based polarity identification contributes to a higher accuracy performance, but induces also a higher computational complexity. Nevertheless, depending on the used methods the reported increase of accuracy of document and sentence classifier range between $2-10\%$ (Pang and Lee, 2004; Wiegand and Klakow,), mostly compared to the baseline (e.g. Naive Bayes) implementations. However, in the majority of cases, only slightly better results could be achieved (Kugatsu Sadamitsu and Yamamoto, 2008; Wiegand and Klakow,). At the focus of almost all approaches, a set of subjectivity terms is needed, either to train a classifier or to extract polarity-related terms following a bootstrapping strategy (Yu and Hatzivassiloglou, 2003).

3 BACKGROUND

3.1 Modeling Opinion Orientation

Following Liu (2010)(Liu, 2010, pp. 5) we formally define an opinion oriented model as follow: A polarity-related document d contains a set of opinion objects $\{o_1, o_2, \ldots, o_q\}$ from a set of opinion holders $\{h_1, h_2, \ldots, h_p\}$. Each opinion object o_j is represented by a finite set of sentiment features, $F = \{f_1, f_2, \ldots, f_n\}$. Each feature $f_i \in F$ is represented in d by a set of term or phrases $W = \{w_{i1}, w_{i2}, \ldots, w_{im}\}$, which correspond to synonyms or associations of f_i and are indicated by a set of feature indicators $I_i = \{i_{i1}, i_{i2}, \ldots, i_{ip}\}$ of the feature. The direct opinion of o_j is expressed through the polarity of the opinion (e.g. positive, negative, neutral) defined as oo_j with respect to the comprised set of features f_j of o_j, the opinion holder h_i and the time or position within the text t_j, an opinion is expressed. The feature indicator i_j reflects thereby the strength of the opinion (e.g. rating scale). Following this definition, contrary opinions within a text document (e.g. phrase or sentence-based) correlate to a (dis-) similarity S of two opinion objects $S(o_j, o_k)$, while a concordance of a polarity is indicated by a high similarity value. At the center of the opinion-oriented model, a mapping from the input document to the corresponding sentiment features with associated indicators ($W \mapsto F$) needs to be established. Meaning, an external resource is needed that embodies not only a set of term or phrase features, but also incorporates the polarity orientation at least as a boolean (positive, negative), preferably on a rating scale (positive, negative, neutral). We refer to these resources as *subjectivity dictionaries*. As we use machine learning classifiers, the similarity function $S(o_j, o_k)$ refers to the similarity between the supervised trained SVM-based opinion models (o_j) and the evaluation set of document opinions (o_k).

3.2 Subjectivity Dictionaries

In recent years a variety of approaches in classifying sentiment polarity in texts has been proposed. However, the number of comprised or constructed subjectivity resources are rather limited. In this section, we describe the most widely used subjectivity resources for the English language in more detail.

Adjective Conjunctions. As one of the first, Hatzivassiloglou et al. (1997) proposed a bootstrapping approach on the basis of adjective conjunctions. Thereby, a small set of manually annotated seed words (1,336 adjectives) were used in order to extract a number of 13,426 conjunctions, holding the same semantic orientation i.e. 'and' indicates an agreement of polarity (nice and comfortable) and 'but' indicates disagreement (nice but dirty). Subsequently, a clustering algorithm separated the sum of adjectives into two subsets of different sentiment orientation (positive or negative). This approach follows the notion that a pair of adjectives (e.g. conjunction in a sentence) will most likely have the same orientation (81% of the unmarked member will have the same semantic orientation as the marked member).

WordNet Distance. Maarten et al. (2004) presented an approach measuring the semantic orientation of adjectives on the basis of the linguistic resource *WordNet* (Fellbaum, 1998). A focus was set on graph-related measures on the syntactic category of adjectives. The geodesic distance is used as a measurement to extract not only synonyms but also antonyms. As a reference dataset, the manually constructed list of the General Inquirer (Stone et al., 1966) was used, comprising $1,638$ polarity-rated terms. Since the evaluation focused on the intersection of both resources (General Inquirer vs. *WordNet*), no additional corpus could be gained.

WordNet-Affect. A related approach in building a sentiment resource, Strapparava and Valitutti (2004)(Strapparava and Valitutti, 2004) studied the synset-relations of *WordNet* with respect to their semantic orientation. Following a bootstrapping-strategy, manually classified seed words were used for constructing a list of 'reliable' relations (e.g. antonym, similarity, derived-from, also-see) out of the linguistic resource. The final dataset, *WordNet-Affect*, comprises $2,874$ synsets and $4,787$ words.

Subjectivity Clues. In 2005, Wiebe et al. (2005) presented the most fine-grained polarity resource. Within the Workshop on Multi-Perspective Question

Table 1: The standard deviation (StdDevi) and arithmetic mean (AMean) of subjectivity features by resource, text corpus (Text) and polarity category (Positive, Negative).

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

Answering (2002) the MPQA corpus was manually compiled. This corpus consists of 10,657 sentences comprising 535 documents. 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).

SentiWordNet. Esuli and Sebastiani (2006) introduced a method for the analysis of glosses associated to synsets of the *WordNet* data set. The proposed subjectivity resource *SentiWordNet* thereby assigns for each synset three numerical scores, describing the objective, negative, and positive polarity of interlinked terms. The used method is based on the quantitative analysis of glosses and a vectorial term representation for a semi-supervised synset classification. Overall, *SentiWordNet* comprises 144,308 terms with polarity scores assigned.

SentiSpin. Takamura et al. (2005) proposed an algorithm for extracting the semantic orientation of words using the *Ising Spin Model* (Chandler, 1987, pp. 119). Their approach focused on the construction of a gloss-thesaurus network inducing different semantic relations (e.g. synonyms, antonyms), and enhanced the built dataset with co-occurrence information extracted from a corpus. The construction of the gloss-thesaurus is based on *WordNet*. With respect to the co-occurrence statistics, conjunctive expressions from the Wall Street Journal and Brown corpus were used. The available subjectivity resource offers a number of $88,015$ words for the English language with assigned Part-of-Speech information and a sentiment polarity orientation.

Polarity Enhancement. Waltinger (2009) proposed an approach to term-based polarity enhancement using a social network. His approach focuses on the reinforcement of polarity-related term features with respect to colloquial language. Using the entries of the *SpinModel* dataset as seed words, associated phrase and term definitions were extracted from the *urban dictionary* project. The enhanced subjectivity resource comprises $137,088$ term features for the English language.

4 METHODOLOGY

With respect to the described approaches in the construction of subjectivity dictionaries, we can identify two different branches. The majority of proposals induce the lexical network *WordNet* as a foundation for either extending or extracting polarity-related semantic relations. Therefore, the constructed term set is limited to the number of entries within *WordNet*, comprising up to $144,308$ polarity features. Other approaches, focused on the manual creation of a subjectivity thesaurus by inducing expert knowledge (manually annotated). These costly built resources consist of a rather small set of polarity features, inducing a dictionary size of up to $6,663$ entries. The questions that arise therefore are: How does the different subjectivity resources perform within the same experimental setup of polarity identification? Does the significant difference (quantity) of used polarity features affect the performance of opinion mining applications? Our methodology focuses on the most widely used and freely available subjectivity dictionaries for the task of sentiment-based feature selection.

4.1 SVM-Classification

The method we have used for the polarity classification is 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, 2002a). This supervised classification technique relies on training a set of polarity classifiers, each of them capable of deciding whether the input stream has a positive or negative polarity, $C = \{+1, -1\}$. The SVM predicts a hyperplane, which separates a given set into two divisions with a maximum margin (the largest possible distance) (Joachims, 2002a). We make use of the SVM^{Light} V6.01 implementation (Joachims, 2002b), using *Leave-One-Out* cross-validation, reporting *F1-Measure* as the harmonic mean between *Precision* and *Recall*. The reported *Accuracy* measures are based on a *5-fold cross-validation*. In each case of the SVM-Classifiers, *Linear*- and *RBF-Kernel* were evaluated in a comparative manner.

4.2 Subjectivity-Feature-Selection

Using SVMs for classifying the sentiment orientation, each input text needs to be converted into a vector representation. This vector consists of a set of significant term features representing the associated document. With respect to the opinion-oriented model, this task corresponds to a mapping between subjectivity features from the particular dictionary, and the textual features of the input document. That is, only those features are selected that occur in the subjectivity lexicon. Since the polarity features can consist of single words as well as multi-word expressions, a sliding window is used, when extracting textual data from the input text. As the feature weighting function, we have used the normalized term frequency ($tf_{i,j}$), defined as

$$tf_{i,j} = \frac{f_{i,j}}{\sum_{k=1}^{n} f_{k,j}} \tag{1}$$

where the number of occurrences of feature i in document j is normalized by the total number of features n in j.

While various subjectivity resources have been proposed in recent years, only a few of them are freely available. In this paper, we evaluate the four most widely used and available resources (Table 1):

- Subjectivity Clues (Wiebe et al., 2005)
- SentiSpin (Takamura et al., 2005)
- SentiWordNet (Esuli and Sebastiani, 2006)
- Polarity Enhancement (Waltinger, 2009)

4.3 German Subjectivity Resource

As described in section 3.2, the majority of subjectivity resources are based on the English language. For the German language there is, to the best of our knowledge, no freely polarity-related dictionary available. We therefore constructed two different German subjectivity dictionaries for the German language, which will be freely available to download after the review process. The construction of these dictionaries is based on a semi-supervised translation of existing English polarity term-sets. That is, we automatically translated each polarity feature into the German language, and manually reviewed the translation quality. Polarity values $(-1, 1)$ were inherited from the English dataset. Since a goal of this paper is to evaluate the correlation between the size of subjectivity dictionaries and the accuracy performance, we have built two different German polarity resources. First, a translation of 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*. Second, we translated the dataset of *SentiSpin* (Takamura et al., 2005), comprising $105,561$ polarity features.. We will refer to this resource as the *German SentiSpin* dictionary. Both resources are freely available for research purposes[2].

5 EXPERIMENTS

5.1 Corpora

We have used two different datasets for the experiments. For the English language we conducted the polarity identification classification 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. 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 and star label information but also the name of the reviewers were removed from the documents. All textual data (term features in the document) were passed through a pre-processing component, that is lemmatized and tagged by a PoS-Tagger. The average number of term features of the comprised reviews is 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

[2]The constructed resources can be accessed at: http://hudesktop.hucompute.org/

Table 2: Accuracy results comparing four subjectivity resources and four baseline approaches.

Sentiment-Method	Accuracy
Naive Bayes - unigrams (Pang et al., 2002)	78.7
Maximum Entropy - top 2633 unigrams (Pang et al., 2002)	81.0
SVM - unigrams+bigrams (Pang et al., 2002)	82.7
SVM -unigrams (Pang et al., 2002)	**82.9**
Polarity Enhancement - PDC (without feature enhancement) (Waltinger, 2009)	81.9
Polarity Enhancement - PDC (with feature enhancement) (Waltinger, 2009)	**83.1**
Subjectivity-Clues SVM Linear-Kernel	**84.1**
Subjectivity-Clues SVM RBF-Kernel	83.5
SentiWordNet SVM Linear-Kernel	**83.9**
SentiWordNet SVM RBF-Kernel	82.3
SentiSpin SVM Linear-Kernel	**83.8**
SentiSpin SVM RBF-Kernel	82.5

Star1 and Star2 against Star 4 and Star 5).

5.2 Results

With respect to the English polarity experiment (see Table 3), we have used not only the published accuracy results of (Pang et al., 2002), using the Naive Bayes (NB), the Maximum Entropy (ME) and the N-Gram-based SVM implementation, but also the results of (Waltinger, 2009), a feature-enhanced SVM implementation as corresponding baselines. As Table 2 shows, the smallest resource, Subjectivity Clues, performs best with $acc = 84.1$. However, SentiWordNet ($acc = 83.9$), SentiSpin ($acc = 83.8$) but also the Polarity Enhancement ($acc = 83.1$) dataset used for feature selection, perform almost within the same accuracy. It can be stated that all subjectivity feature selection resources clearly outperform not only the well known NB and ME classifier but also the N-Gram-based SVM implementation. Not surprisingly, with respect to the feature coverage of the used subjectivity resources (see Table 1), we can argue that the size of the dictionary clearly correlates to the coverage (arithmetic mean of polarity-features selected varies between $76.83 - 241.36$). Interestingly, the biggest dictionary with the highest coverage property does not outperform the resource with the lowest number of polarity-features. In contrast, we can state that operating in the present settings, on $6,663$ term features (in contrast to $144,308$ of *SentiWordNet*), seem to be a sufficient number for the task of document-based polarity identification. This claim is also supported by the evaluation F1-Measure results as shown in Table 3. All subjectivity resources nearly perform equally well (F1-Measure results range between $82.9 - 83.9$). In this $Leave-One-Out$ estimation, the polarity-enhanced implementation performs with a touch better than the other resources.

Table 4 shows the results of the new build German subjectivity resources, used for the document-based polarity identification. With respect to the correlation of subjectivity dictionary size and classification performance, similar results can be achieved. 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. In general, in terms of Kernel-Methods, we can argue that RBF-Kernel are inferior to the Linear-Kernel SVM implementation, though only to a minor extend. With reference to the coverage of subjectivity dictionaries for a polarity-based feature selection - *size does matter*. However, the classification accuracy results indicate - for both languages - that a smaller but controlled dictionary contributes to the accuracy performance (almost equally to big-sized data) of opinion mining systems.

6 CONCLUSIONS

This paper proposed an empirical study to machine learning-based sentiment analysis. We systematically analyzed the four most widely used subjectivity resources for the task of sentiment polarity identification. The evaluation results showed that the size of subjectivity dictionaries does not correlate with classification accuracy. Smaller but more controlled dictionaries used for a sentiment feature selection perform within a SVM-based classification setup equally good compared to the biggest available resources. We

Table 3: F1-Measure evaluation results of an English subjectivity feature selection using SVM.

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

Table 4: F1-Measure evaluation results of a German subjectivity feature selection using SVM.

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

can conclude, that combining a polarity-based feature selection with machine learning, SVMs using Linear-Kernel exhibit the best performance ($acc = 84.1, f1 = 83.9$). In addition, we proposed a new freely available German subjectivity resource, which was evaluated using a product review corpus. The results of the German polarity identification experiments, with an F1-Measure of 85.9 are quite promising.

ACKNOWLEDGEMENTS

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

REFERENCES

Agarwal, A., Biadsy, F., and McKeown, K. (2009). Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In *EACL2009*, Athens, Greece.

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

Becker-Asano, C. and Wachsmuth, I. (2009). Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*.

Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*. Oxford University Press.

Chaovalit, P. and Zhou, L. (2005). Movie review mining: a comparison between supervised and unsupervised classification approaches. *Hawaii International Conference on System Sciences*, 4:112c.

Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *WWW '03: Proceedings of the twelfth international conference on World Wide Web*, pages 519–528. ACM Press.

Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC06*, pages 417–422.

Fellbaum, C., editor (1998). *WordNet. An Electronic Lexical Database*. The MIT Press.

Hatzivassiloglou, V. and McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 174–181, Morristown, NJ, USA. Association for Computational Linguistics.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *KDD '04: Proceedings of the tenth*

ACM SIGKDD international conference on Knowledge discovery and data mining, pages 168–177, New York, NY, USA. ACM.

Joachims, T. (2002a). *Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA.

Joachims, T. (2002b). SVM light, http://svmlight.joachims.org.

Kennedy, A. and Inkpen, D. (2006). Sentiment classification of movie reviews using contextual valence shifters. *Computational Intelligence*, 22(2):110–125.

Kugatsu Sadamitsu, S. S. and Yamamoto, M. (2008). Sentiment analysis based on probabilistic models using inter-sentence information. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, B. M. J. M. J. O. S. P. D. T., editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, 2:568.

Maarten, J. K., Marx, M., Mokken, R. J., and Rijke, M. D. (2004). Using wordnet to measure semantic orientations of adjectives. In *National Institute for*, pages 1115–1118.

Mehler, A., Geibel, P., and Pustylnikov, O. (2007). Structural classifiers of text types: Towards a novel model of text representation. *Journal for Language Technology and Computational Linguistics (JLCL)*, 22(2):51–66.

Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In Lin, D. and Wu, D., editors, *Proceedings of EMNLP 2004*, pages 412–418, Barcelona, Spain. Association for Computational Linguistics.

Pang and Lee (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *In Proceedings of the ACL*, pages 271–278.

Pang, B. and Lee, L. (2005). Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 115–124, Morristown, NJ, USA. Association for Computational Linguistics.

Pang, B. and Lee, L. (2008). *Opinion Mining and Sentiment Analysis*. Now Publishers Inc.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79–86, Morristown, NJ, USA. Association for Computational Linguistics.

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

Stone, P. J., Dunphy, D. C., Smith, M. S., and Ogilvie, D. M. (1966). *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

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

Taboada, M., Brooke, J., and Stede, M. (2009). Genre-based paragraph classification for sentiment analysis. In *Proceedings of the SIGDIAL 2009 Conference*, pages 62–70, London, UK. Association for Computational Linguistics.

Takamura, H., Inui, T., and Okumura, M. (2005). Extracting semantic orientations of words using spin model. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 133–140, Morristown, NJ, USA. Association for Computational Linguistics.

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

Turney, P. D. (2001). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 417–424, Morristown, NJ, USA. Association for Computational Linguistics.

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

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

Wiebe, J. and Riloff, E. (2005). Creating subjective and objective sentence classifiers from unannotated texts. In *Proceeding of CICLing-05, International Conference on Intelligent Text Processing and Computational Linguistics.*, volume 3406 of *Lecture Notes in Computer Science*, pages 475–486, Mexico City, MX. Springer-Verlag.

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

Wiegand, M. and Klakow, D. The role of knowledge-based features in polarity classification at sentence level.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354, Morristown, NJ, USA. Association for Computational Linguistics.

Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of EMNLP'03*.