SentiBA: Lexicon-based Sentiment Analysis on German Product Reviews

Markus Dollmann
Heinz Nixdorf Institut
Universität Paderborn
Fürstenallee 11
33102 Paderborn
dollmann@mail.upb.de

Michaela Geierhos
Heinz Nixdorf Institut
Universität Paderborn
Fürstenallee 11
33102 Paderborn
geierhos@hni.upb.de

Abstract

In this paper, we describe our system developed for the GErman SenTiment AnaLysis shared Task (GESTALT) for participation in the Maintask 2: Subjective Phrase and Aspect Extraction from Product Reviews. We present a tool, which identifies subjective and aspect phrases in German product reviews. For the recognition of subjective phrases, we pursue a lexicon-based approach. For the extraction of aspect phrases from the reviews, we consider two possible ways: Besides the subjectivity and aspect look-up, we also implemented a method to establish which subjective phrase belongs to which aspect. The system achieves better results for the recognition of aspect phrases than for the subjective identification.

1 Introduction

The Maintask 2 aims at extracting aspects and subjective phrases and their relation in German product reviews (Ruppenhofer et al., 2014).

The system implementation for this shared task is based on previous unpublished work. The original goal was to use linguistic phenomena in order to determine the contextual polarity of subjective phrases for the sentiment classification of reviews at the document level. The implementation, called SentiBA, takes the three polarity classes positive, neutral and negative into account. It considers contextual valence shifter such as negation, intensifiers, modals, questions and a few rules for

irony detection. The consideration of these contextual valence shifters had a great impact on the performance of the sentiment analysis task.

For GESTALT, we extended and improved the functionality of SentiBA by including aspect identification and by optimizing the recognition of subjective (polarity) words and phrases. Furthermore, we also implemented a mapping of subjective expressions to their target aspect phrases.

This paper is organized as follows: In Section 2, we sum up related work. In Section 3, the lexical resources are introduced. Section 4 provides a conceptual overview of our approach for this shared task. In Section 5, we present the results of our system obtained on the evaluation data and explain the different run settings, followed by a short discussion and conclusion in Section 6.

2 Related Work

Sentence or aspect-based sentiment analysis usually consists of two steps: First identify and then classify subjective expressions into positive and negative terms. For this task, only the subjectivity classification is of interest. Different methods have been developed to recognize subjective sentences. A common technique, the lexicon-based approach, uses lists of opinion words (e.g. Ding et al. (2008)). If a sentence contains one or more words of that list, it is assumed to be subjective. Another common approach uses machine learning techniques to extract subjective phrases by previously learned patterns. Our implementation is inspired by lexicon-based approaches, to match the subjective expressions in sentences more easily and to deal with linguistic phenomena such as valence shifters.

Valence shifters (Polanyi und Zaenen, 2004) are words and phrases that can shift or change semantic orientation. Although we ignore the semantic orientation of words and phrases for this task, we have to consider some of these valence shifters, too. Since valence shifters have an impact on subjective expressions, they should be stored together.

We identified two rules to find additional subjective expressions which are not covered by the sentiment lexicon. One of these rules introduced by Hatzivassiloglou und McKeown (1997) deals with the conjunction "and". It says that conjoined adjectives usually have the same orientation. In the sentence "This car is beautiful and spacious" where "beautiful" is known to be subjective, it can be inferred that "spacious" is also subjective. Further if "beautiful" is known to be positive, "spacious" is very likely to be also positive, because people usually express the same sentiment on both sides of a conjunction (Liu, 2012). A similar rule is about the connective "but" which is similar to the rule explained above, but has a contrary impact on the polarity of the words (Hatzivassiloglou und McKeown, 1997).

Hu und Liu (2004) present a frequency-based approach to identify aspect phrases. Nouns that are frequently used are likely to be true aspects (called frequent aspects). When different reviewers tell different (irrelevant) stories, the words used to discuss the product aspects/features converge. These words are the main aspects.

3 Resources

To identify subjective expressions, we used the sentiment lexicon SentiWS, which contains 1,650 positive and 1,818 negative word lemmas, which sum up to 15,649 positive and 15,632 negative word forms incl. their inflections (Remus et al., 2010). We also used a list of negation words and intensifiers, which were optained from the German version of SentiStrength[1].

The USAGE data set serves as training data for this shared task (Klinger und Cimiano, 2014). The data set contains annotations for more than 600 German Amazon reviews covering six different domains: Coffee machines, cutlery sets, microwaves, toasters, trash cans, vacuum cleaners and washing machines. We divided the training set into two parts. The coffee machine reviews were used to test our system. The other reviews were used to generate blacklists of subjective and aspect phrases, by counting for all expressions, how often they were correctly or incorrectly identified (see Sections 4.2 and 4.3).

We also created a subjectivity lexicon from the annotated training data provided for this maintask (except the coffee machine reviews). In the following we will call this lexicon the USAGE lexicon. We extracted all subjective words and phrases from the training data, counted the number of occurrence for each expression and created a frequency list. We tested the USAGE lexicon in conjunction with SentiWS on the coffee machine reviews and achieved better results than with SentiWS alone. Due to misidentifications in different domains, we decided to manually delete domain-dependent expressions, by the estimation of the authors. We received a list of subjective words and phrases that is domain independent and contains typical expressions used in product reviews, like "5-stars" or "strong buy recommendation". Due to these adaptations, we achieved even better results in our tests. The created USAGE lexicon contains 13 subjective words and 267 subjective phrases.

4 Implementation

In this section, we present our implementation design. Figure 1 gives an overview of the sequential steps and the required resources. These steps will be described in this section (see Sections 4.1-4.5). First, SentiBA preprocesses each product review. Subsequently the tool identifies subjective and aspect phrases. Then SentiBA indicates corresponding subjective phrases for each aspect phrase. Finally, all collected information is stored in a structured format.

4.1 Preprocessing

Before identifying subjective and aspect phrases, we preprocess each review by means of the Apache OpenNLP toolkit[2].

---

[1] http://www.ofai.at/research/
interact/resources/SentiStrength_DE/
download_form.html
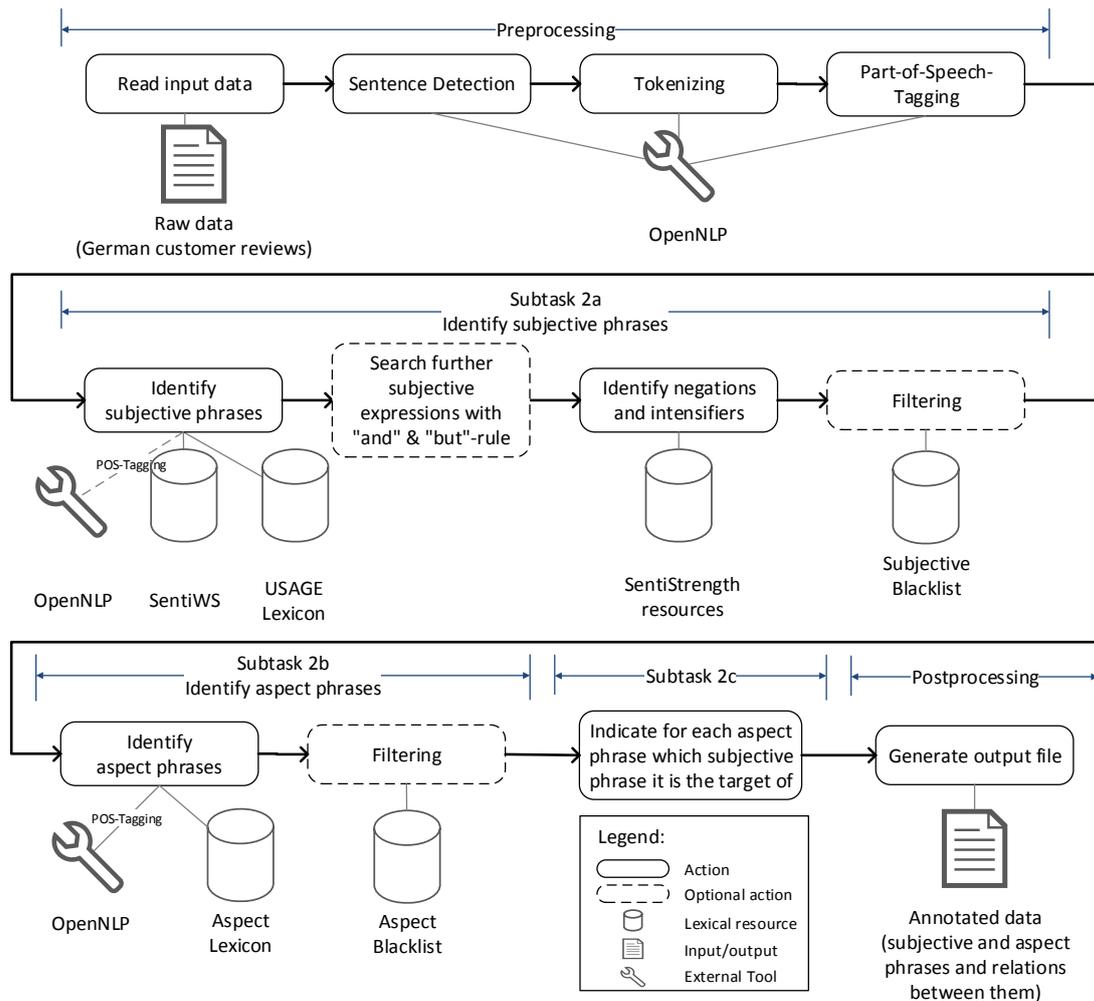
[2] https://opennlp.apache.org

Figure 1: System overview: Steps and resource usage

We used the Sentence Detector (trained on TIGER data) from OpenNLP to split the reviews in single sentences. After that, they were tokenized by the OpenNLP Tokenizer (trained on TIGER corpus). The data structure allows us to add individual tags to every token. That way, we label tokens as subjective, aspect, negation, intensifier or any other predefined tag using the OpenNLP POS-Tagger (maxent model trained on TIGER corpus).

4.2 Subtask 2a: Identify subjective phrases

As already mentioned, we extended SentiBA by adding the sentiment lexicon SentiWS to process German reviews. We also improved the identification of subjective (polarity) words and

phrases in different ways, independently from the research goal of our previous work.

To identify subjective words, SentiBA looks up every word of a review in the sentiment lexicon SentiWS. If the word exists in SentiWS, it will be annotated as subjective. When POS-Tagging is enabled, the word is only labeled as subjective if also the POS tag of the word in the review is equal to its POS tag in the lexicon. Additionally SentiBA also checks every word and phrase, in the USAGE lexicon. In this case POS tags are not considered any more.

To extend the recognized subjective words to subjective phrases, we identify negation words and intensifiers by a single token comparison with a list of negation words or intensifiers. In this

case, we add a specific tag to these words. Since we are interested in subjective phrases so far and not in the polarity of these phrases, a further processing is not necessary. In the postprocessing step (see Section 4.5) these identified negations and intensifiers will be combined with the subjective words to become phrases.

We also detect additional subjective words (which are not included in SentiWS) by using patterns with the conjunctions "and" (in German: „und") and the connective "but" (in German: „aber"). If a sentence contains the word „und" or „aber", SentiBA searches in the left and right context of the target word within a given window. If an already identified subjective word is found, SentiBA looks in the other direction of the sentence, for a given distance from the words „und" or „aber" for an unidentified adjective. In our tests, the best performance was achieved by a word distance of one, which means that the adjective and the already identified word are directly next to the word „und" or „aber". If SentiBA locates an adjective, it will label it as a subjective word. To filter common misidentified subjective

Aspect	Translation	#Incorrect	#Correct
leider	sadly	55	0
gut	good	36	57
einfach	easily	28	19
alten	old	22	0
schnell	fast	22	20
alte	old	20	0
kleine	small	18	0
neue	new	18	0
genau	exactly	16	0
wieder kaufen	buy again	15	0

Table 1: 10 most frequent misidentified subjective words and phrases

expressions, we created a blacklist. To generate this blacklist, we counted for all identified subjec-

tive words and phrases from the training data (except the coffee machine reviews) how often they were correctly or incorrectly identified. Table 1 shows the most frequent misidentified subjective expressions together with their corresponding frequency of being (in)correctly identified.

4.3 Subtask 2b: Identify aspect phrases

We implemented two different approaches to identify aspect phrases in product reviews: A frequency-based approach and a naive approach, which nevertheless achieves better results.

Frequency-based approach

One approach was to identify aspect phrases through an aspect lexicon, which contains the most frequent candidates for aspect phrases from product reviews for the specific domain. We identified potential aspects by noun POS tags. The 10 most frequent potential aspects for the domain "coffee machine" are given in Table 2. We gen-

Aspect	Translation	Frequency
Kaffee	coffee	90
Maschine	machine	71
Kaffeemaschine	coffee machine	67
Kanne	pot	35
Wasser	water	20
Preis	price	13
Gerät	device	12
Thermoskanne	thermos	11
Tassen	mugs	11

Table 2: 10 most frequent aspect candidates for coffee machines

erated a frequency list for all potential aspect expressions. To identify aspects, we look up each word or phrase in that aspect lexicon, under the assumption that a specific threshold is exceeded. Surprisingly, starting by a threshold of one, the higher the threshold the lower the F-Score for the aspect identification. While the precision increases with a higher threshold, the recall drops very quickly. Our second approach achieved considerable better results.

Does each noun describe an aspect?

The more satisfying approach is also based on the POS tag for nouns. Instead of the frequency-based approach, SentiBA now assumes that every noun in the product review represents an aspect. Just like in the subjectivity identification, we created a blacklist to filter common misidentified expressions. To generate this blacklist, we counted for all identified nouns (and noun phrases) from the training data (except the coffee machine reviews) how often they were correctly or incorrectly identified. Table 3 shows the most frequent misidentified aspects together with their corresponding frequency of being (in)correctly identified. This very simple approach achieves remarkably better results in our tests on the coffee machine reviews.

Aspect	Translation	#Incorrect	#Correct
Zeit	time	36	24
Jahre	years	27	0
Jahr	year	26	0
Gebrauch	use	23	4
Für	for	22	0
Jahren	years	22	0
Probleme	problems	21	0
Fazit	conclusion	21	0
Problem	problem	18	0
Tag	day	17	0

Table 3: 10 most frequent misidentified aspects

4.4 Subtask 2c: Indicate for each aspect phrase which subjective phrase it is the target of

We applied a quite simple approach to indicate corresponding subjective phrases for each aspect phrase. SentiBA calculates for each identified aspect phrase from Subtask 2b the token distance to every identified subjective phrase, which is in the same sentence as the aspect phrase. The subjective phrase with the shortest distance to the aspect

phrase will be taken as the subjective expression for that aspect phrase.

This approach can easily be extended in future by adding multiple subjective phrases to aspects, e.g. if multiple subjective phrases in the same sentence are connected by words like "and" or "but". Moreover, coreference resolution is not considered in this approach. A possible attempt could be to search backward for the next aspect phrase and match the coreference word with this aspect.

4.5 Postprocessing

In the postprocessing step SentiBA stores all previously collected information into two output files: One file for the identified subjective and aspect phrases and one file for the relations between them.

SentiBA saves every word of the input review, which was tagged as subjective in the output file. Therefore SentiBA links the neighboring subjective words to phrases and also adds neighboring negations and intensifiers to these words or phrases. It is done in a similar way for the identified aspect words, while neighboring aspect words are saved as an aspect phrase. Additionally the identified relations from Subtask 2c are stored in the relation file.

5 Results

SentiBA was tested with different settings. Because of the poor results during our own tests, we decided to drop the frequency-based aspect identification approach and only pursued the approach presupposing each noun as an aspect.

We devided our evaluation runs as shown in Table 4. In three of five runs we used the subjective

Run	Blacklists	POS-Tagging	"and"& "but"-rule
1	√	X	X
2	√	X	√
3	√	√	√
4	X	X	X
5	X	√	X

Table 4: Settings for the different runs

	Precision	Recall	F_1
Run 1			
Subtask 2a	**0.527**	0.312	0.392
Subtask 2b	**0.555**	0.622	**0.587**
Subtask 2c	**0.126**	0.138	**0.132**
Run 2			
Subtask 2a	0.516	0.320	0.395
Subtask 2b	**0.555**	0.622	**0.587**
Subtask 2c	0.124	0.138	0.131
Run 3			
Subtask 2a	0.503	0.260	0.342
Subtask 2b	0.530	0.614	0.569
Subtask 2c	0.118	0.117	0.118
Run 4			
Subtask 2a	0.443	0.359	0.396
Subtask 2b	0.477	**0.650**	0.550
Subtask 2c	0.095	**0.148**	0.116
Run 5			
Subtask 2a	0.432	**0.367**	**0.397**
Subtask 2b	0.477	**0.650**	0.550
Subtask 2c	0.092	0.143	0.112

Table 5: Results from the different runs on the test data

and aspect blacklists to filter common misidentified subjective and aspect expressions. Although these blacklists had a positive influence during our tests on the coffee machines, we decided to also perform runs without these blacklists, if the main aspect or subjective words and phrases of the new category are part of these blacklists. We also decided to have runs with and without POS-Tagging. POS-Tagging helps to identify different word senses, but also decreases the number of recognitions in the lexicon. The last difference in the runs is the application of rules to identify new subjective words by usage of the conjunction "and" and the connective "but".

We decided to have runs in- and excluding these rules, in order to examine whether new subjective words can be identified with this method.

But the error rate should not be underestimated.

The results from the different runs on the test data are given in Table 5. The best results for identifying subjective phrases (see F-Score in Subtask 2a) were achieved by run no. 5, where the subjective blacklist was not used, POS-Tagging was enabled and the both conjunction-rules were disabled. The usage of POS-Tagging improves the recall, but decreases the precision (compare with run no. 4). The usage of the subjective blacklist increases the precision remarkably, but decreases the recall seriously.

The best results for identifying aspect phrases (see F-Score in Subtask 2b) were achieved by the runs no. 1 and no. 2, when the aspect blacklist was used and POS-Tagging was disabled. The usage of the "and" & "but"-rules had no impact on the aspect identification.

The results for the matching of aspect phrases to subjective phrases depend on the results of Subtask 2a and 2b. The best result was delivered by run no. 1, where also the aspect identification achieved the best result.

In comparison to our own evaluation on the coffee machine reviews (see Table 6) the results on the test data are poorer. The best F-Score reached on the test data by identifying subjective phrases is 0.397, on the coffee machine reviews the score is 0.453. For identifying aspect phrases, the best F-Score on the test data is 0.587, while on the coffee machine reviews it is 0.634.

	Run 1	Run 2	Run 3	Run 4	Run 5
Subtask 2a	**0.453**	0.452	0.366	0.431	0.359
Subtask 2b	0.663	0.663	**0.634**	0.620	0.595
Subtask 2c	**0.199**	0.195	0.158	0.168	0.135

Table 6: F-Scores from runs on coffee machine reviews from training data (Annotator 1)

SentiBA achieves an F-Score of 0.132 on the test data for matching aspect phrases with subjective expressions, while it achieves on the coffee machine reviews a score of 0.199. This shows, that SentiBA together with the sentiment lexicon SentiWS is highly domain sensitive.

6 Conclusion and Future Work

We presented a system for subjective phrase and aspect extraction from product reviews. We pursued a lexicon-based approach using SentiWS and a newly created and manually edited subjective lexicon from the training data. To identify aspect phrases, we implemented two approaches: A frequency-based approach, which identifies aspect phrases through an aspect lexicon that contains the most frequent candidates for aspect phrases and an even more satisfying approach based only on the noun POS tag, where our system assumes that every noun in the product review represents an aspect. We also conducted a simple matching method that assigns each aspect phrase to its corresponding subjective phrase. While the system achieves satisfactory results in the recognition of aspect phrases, the subjective identification and especially the matching should be improved in further work. The comparison between the results from the test data and the results from an excluded part of the training data showed that our implementation is highly domain sensitive. Moreover it shows that the different run settings in various domains have varying results. The frequent nouns approach for identifying aspect phrases gave poor results on the test data; so it was not used in the test runs. In future work, this approach could be improved by searching frequent nouns on a bigger training corpus or by searching for more reviews from the same domain in the Internet. The matching of aspect and subjective phrases could be improved by applying coreference resolution and by further research for better rules to indicate which subjective phrase belongs to which aspect phrase.

References

Xiaowen Ding, Bing Liu, und Philip S. Yu. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, WSDM '08, S. 231–240, New York, NY, USA. ACM.

Vasileios Hatzivassiloglou und Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*, ACL '98, S. 174–181, Stroudsburg, PA, USA. Association for Computational Linguistics.

Minqing Hu und Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, S. 168–177, New York, NY, USA. ACM.

Roman Klinger und Philipp Cimiano. 2014. The usage review corpus for fine grained multi lingual opinion analysis. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, und Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may. European Language Resources Association (ELRA).

B. Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis digital library of engineering and computer science. Morgan & Claypool.

Livia Polanyi und Annie Zaenen. 2004. Contextual valence shifters. In *Working Notes — Exploring Attitude and Affect in Text: Theories and Applications (AAAI Spring Symposium Series)*.

R. Remus, U. Quasthoff, und G. Heyer. 2010. Sentiws – a publicly available german-language resource for sentiment analysis. In *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*, S. 1168–1171.

Josef Ruppenhofer, Roman Klinger, Julia Maria Struß, Jonathan Sonntag, und Michael Wiegand. 2014. Iggsa shared tasks on german sentiment analysis (gestalt). In Gertrud Faaßund Josef Ruppenhofer, editors, *Workshop Proceedings of the 12th Edition of the KONVENS Conference*, Hildesheim, Germany, October. Universität Hildesheim.