Evaluation of an Algorithm for Aspect-Based Opinion Mining Using a Lexicon-Based Approach

Florian Wogenstein
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
fwogenstein@iisys.de

Johannes Drescher
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
jdrescher@iisys.de

Dirk Reinel
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
dreinel@iisys.de

Sven Rill
Goethe University Frankfurt
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
srill@iisys.de

Jörg Scheidt
University of Applied Sciences
Alfons-Goppel-Platz 1
95028 Hof, Germany
jscheidt@iisys.de

ABSTRACT

In this paper, we present a study of aspect-based opinion mining using a lexicon-based approach. We use a phrase-based opinion lexicon for the German language to investigate, how good strong positive and strong negative expressions of opinions, concerning products and services in the insurance domain, can be detected. We perform experiments on hand-tagged statements expressing opinions retrieved from the Ciao platform. The initial corpus contained about 14,000 sentences from 1,600 reviews. For both, positive and negative statements, more than 100 sentences were tagged. We show, that the algorithm can reach an accuracy of 62.2% for positive, but only 14.8% for negative utterances of opinions. We examine the cases, in which the opinion could not correctly be detected or in which the linking between the opinion statement and the aspect fails. Especially, the large gap in accuracy between positive and negative utterances is analysed.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*; I.2.7 [**Natural Language Processing**]: Text analysis

General Terms

Algorithms

Keywords

Aspect-based opinion mining, lexicon based approach, opinion lexicon

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
WISDOM '13, August 11 2013, Chicago, USA.
Copyright 2013 ACM 978-1-4503-2332-1/13/08 ...$15.00.

1. INTRODUCTION

The need of techniques for an automatic analysis of textual data has raised as the amount of text data has increased in the Web 2.0. Beside other issues, the extraction of opinions from text data becomes more and more important.

Opinion mining can be performed on several levels. The document-level opinion classification is the less granular approach. The aspect-based opinion mining, aiming to find out opinions uttered about aspects or features of entities, is the most fine-grained approach.

An example for entities are technical devices like mobile phones with aspects such as the display or the battery. Another example for an entity is a company with its products or services as aspects. Even human beings can be regarded as entities about which opinions are uttered. In this case, aspects are the attributes describing them or, e.g., their skills on specific things.

We describe a quite simple algorithm used to extract aspects and opinion bearing phrases, retrieve opinion values from an opinion lexicon and map the phrases to the aspects. The opinion lexicon lists opinion phrases and their opinion values for the German language. Thus, as the phrases directly include negation words and valence shifters, the task of opinion composition is much easier compared with other approaches, where the opinion values have to be derived from single word opinion lexicons.

We test our algorithm using contributions in the area of insurances retrieved from a German review platform. Thus, we deal with opinion utterances concerning insurance companies, their products and services.

2. RELATED WORK

During the last decade, a lot of research work has been done in the area of opinion mining.

Overviews of the different topics of opinion mining or sentiment detection have been given in [22] and recently in [12] as well as in [6].

The aspect-based opinion mining can be performed using supervised learning techniques, several groups have discussed this approach, see [2, 14, 40]. However, the supervised learning approach is highly dependent on the training data used.

As the training data normally covers one specific domain, the trained system is not easily applicable to a wider range of application domains.
The lexicon-based approach has been applied and discussed in several publications, see for example [9]. Here, opinion lexicons are used to determine the opinion values for statements expressing opinions. Most opinion lexicons list words with their opinion values, negation words and valence shifters (intensifiers and reducers). Thus, the opinion values of phrases have to be composited by the several values for the basic words. The procedure of opinion composition is discussed in [7, 18, 23, 25, 36].
Opinion lexicons exist for several languages. For the English language, e.g., widely used resources are SentiWordNet [1, 10, 11], WordNet-Affect [34] and Semantic Orientations of Words [35], all three generated using the WordNet®[24] lexical database, the Subjectivity Lexicon [41], SenticNet [5] and lists of positive and negative opinion words provided by [21]. Also for the German language, opinion lexicons exist, namely a polarity lexicon described in [8], listing about 8,000 opinion words with their opinion values, GermanPolarityClues [39] with more than 10,000 opinion words and SentiWS [30] with about 3,500 words. All these resources only include opinion values for single words.
The generic approach to derive the opinion lexicon used for this work has been described in [32], the generation of the list for the German language, called Sentiment Phrase List (SePL)[1] has been discussed in [31]. It includes 2,833 phrases with a length of up to five words each. However, it just contains adjective- and noun-based phrases, but it does not yet include any verbs or verb-based phrases.
Also opinion lexicons for other languages exist, e.g., for Spanish [4].
Applications of opinion mining are widely discussed. Online reviews are used for several purposes, examples are classification [28, 38], summarization [27] and the evaluation of the helpfulness of reviews [26]. Another emerging field is the detection of review spam, see [15, 16].

3. THE ALGORITHM

The algorithm to perform the aspect-based opinion mining is done in several steps which are described in the Sections 3.1 to 3.4. Figure 1 depicts the whole process.
At the beginning, some preprocessing steps are necessary. Afterwards, the aspects, which are relevant for the domain under consideration, are extracted. The next step is the detection and classification of opinion bearing phrases. An opinion lexicon for the German language is used to classify these phrases into strong positive and strong negative expressions of opinions. At the end, the opinion bearing phrases are linked to the associated aspects.
The study does not aim to obtain results for specific insurance companies. Thus, in our examples, the names of the insurances are masked as ABC and XYZ in the following.

3.1 Preprocessing

In a first step, sentences are separated using the Apache OpenNLP[2] Sentence Detector. Afterwards, both, the Apache OpenNLP Tokenizer and the Apache OpenNLP Part-Of-Speech Tagger, are applied to separate words and to assign

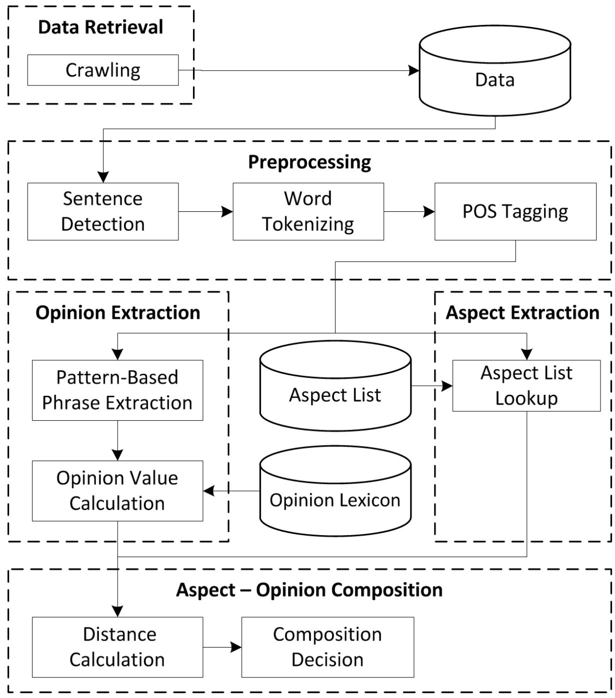

Figure 1: Overview of the algorithm.

the correct POS-tags to them. Therefore, the model corresponding files for the German language are used. For the POS-tagger the maximum entropy model, trained using the TIGER treebank [3], was used. The POS-tags are given in the Stuttgart-Tübingen Tagset (STTS) [33] systematics.
Sometimes, the POS-tagging is erroneous. Especially in cases, where an adjective is the first word in a sentence and therefore written with a capital letter, it occasionally gets tagged as a noun. To deal with this problem, all words at the beginning of a sentence tagged as nouns are tagged again with the capital letter replaced by a small letter. In addition, the Stanford POS-tagger [37] is used. If both algorithms tag the word as an adjective, the POS-tag is changed from noun to adjective.

3.2 Aspect Definition and Extraction

As we have chosen the insurance domain for our experiments, only entities and aspects from this domain have to be extracted from the text. Therefore, an entity and aspect model was produced in order to organize the entities (insurances) and connected aspects.
This model was generated in a semi-automated way. In a first step, entities and aspects were collected manually in a base list. Afterwards, this list was extended using the community-generated German synonym lexicon OpenThesaurus [3]. At the end, the list was normalized by lemmatizing all words using the German Morphology-Lexicon[4], which is based on Morphy[5] [19].
In the following, we do not distinguish between entities and aspects any more, as they are treated exactly in the same

[1]http://www.opinion-mining.org/
[2]http://opennlp.apache.org/
[3]http://www.openthesaurus.de/
[4]http://www.danielnaber.de/morphologie/
[5]http://www.wolfganglezius.de/doku.php?id=cl:morphy

way. Table 1 shows some entities and aspect groups as well as examples for synonyms and sub-aspects.

Entities	Synonyms
ABC-Versicherung - *ABC Insurance*	ABC - *ABC* ABC-Versicherung - *ABC-insurance*
XYZ AG - *XYZ plc*	XYZ - *XYZ* XYZ-Versicherung, XYZ Versicherung - *XYZ insurance* XYZ-AG - *XYZ-plc*
Aspect Groups	**Examples for Sub-Aspects**
Produkt - *product*	KFZ-V. - *automobile insurance* Sach-V. - *property insurance* Unfall-V. - *accident insurance*
Kosten - *costs*	Versicherungsbeitrag - *insurance premium* Gebühr - *fee*, Rabatt - *discount*
Service - *service*	Servicemitarbeiter - *service staff* Kundenservice - *customer service* Beratung - *consulting*
Konditionen - *conditions*	Vertrag - *contract* Angebot - *offer* Tarif - *tarif*

Table 1: Some entities and aspect groups in the insurance domain.

The extraction of the aspects is carried out as a simple search. Due to the fact, that some aspects span over more words, the longest possible aspect phrase is taken, e.g., as "Dienst" - *'service'* and "Öffentlicher Dienst" - *'public service'* both are aspects, the latter will be taken as the aspect for the opinion mapping.
Within the extraction of the aspects from the texts, the same lemmatizer as used for the generation of the list is applied.

3.3 Handling of Opinion Bearing Phrases

In our approach, an opinion bearing phrase consists of at least one opinion bearing adjective or noun and up to four additional parts like intensifiers, negation words, adverbs or other adjectives. Examples for opinion bearing phrases are "brilliant" - *'brilliant'*, "sehr gut" - *'very good'*, "nicht wirklich besonders gut" - *'not really especially good'* or "Schrott" - *'junk'*. Two steps have to be performed to obtain the opinion expressed. In a first step, the opinion bearing phrases are extracted based on the patterns mentioned above. Afterwards, opinion values are looked up from the opinion lexicon and converted into three classes (strong negative, strong positive, other).
As the opinion lexicon we chose the phrased-based opinion list for the German language SePL, described in [31].

3.3.1 Extraction of the Phrases

For extracting the opinion bearing phrases, the same patterns as for the generation of the opinion lexicon were used [31]. Due to the fact, that in the opinion lexicon all words are lemmatized, a lemmatizing of the phrases was necessary to obtain the opinion values from the lexicon. Some minor changes had to be applied to the procedure in order to lemmatize all words in the right way and especially to suppress the lemmatizing for comparative and superlative forms. Otherwise, phrases like "beste" - *'best'* would have been lemmatized to "gut" - *'good'* which is not wanted as the former expresses a stronger opinion compared to the latter.

3.3.2 Application of the Opinion Lexicon

The opinion lexicon includes opinion values for both, single opinion bearing words and phrases with a length of up to five words.
While using the opinion lexicon, in most of the cases it is possible to obtain an opinion value for a given opinion bearing phrase extracted from the text directly. An example for this is the phrase "wirklich sehr gut" - *'really very good'*. However, sometimes phrases are missing in the opinion lexicon. If the phrase just consists of one word, there is no possibility to determine an opinion value. If the phrase consists of more than one word, the phrase is shortened by one word and another lookup in the opinion lexicon is performed. If the shortened phrase occurs in the list, the corresponding opinion value is taken in all cases where the shortening does not cut off negation words. If, for example, the phrase "sehr sehr gut" - *'very very good'* occurs but is not present in the opinion lexicon, the phrase is shortened to "sehr gut" - *'very good'*. On the other hand, if the phrase "nicht ausgesprochen gut" - *'not exceptionally good'* is found and its opinion value can not retrieved from the opinion lexicon, it will not be shortened as omitting the negation word would change the tonality drastically.
At the end, each opinion phrase is categorized as strong positive, if the opinion value is greater than 0.67 and as strong negative, if the opinion value found is smaller than -0.67. At this point, we assume that opinion values can be classified into three equal-sized intervals (negative, neutral and positive) and that both, the positive and negative class, can be divided again into subclasses representing strong and weak polarities.
Table 2 lists examples of strong positive and strong negative opinion bearing phrases with their opinion values. We distinguish between adjective- and noun-based phrases. All examples are directly taken from the test data described in Section 4. Also shown is one example of an opinion phrase which has an opinion value of less than 0.67, thus being not regarded as strong positive ("freundlich" - *'kind'*).

Adjective-Based Phrases	*OV*	$\|sp/sn$
großartig - *great*	0.94	| sp
sehr günstig - *very low priced*	0.89	| sp
kompetent - *competent*	0.77	| sp
freundlich - *kind*	0.58	| –
mies - *lousy*	-0.71	| sn
nur schlecht - *just bad*	-0.88	| sn
Noun-Based Phrases	***OV***	$\|sp/sn$
Herz - *heart*	0.83	| sp
Mogelpackung - *bluff package*	-0.70	| sn
Frechheit - *impertinence*	-0.91	| sn

Table 2: Some words and phrases with their opinion values and the categorization into strong positive (sp) and strong negative (sn).

3.4 Distance-Based Linking

The linking of the opinion phrases to the aspects is done using a distance-based approach applied on the sentence-level. All strong positive or negative opinion phrases are linked to the next aspect found in a sentence according to the word position.

A simple example is the sentence "Ich bin sehr enttäuscht von dem Service." - *'I am very disappointed of the service.'*. The opinion phrase "sehr enttäuscht" - *'very disappointed'*, having an opinion value of -0.78 and thus being strong negative, is linked to "Service" - *'service'*. The result is an opinion tuple giving the opinion phrase, the tonality (sn = strong negative, sp = strong positive) and the aspect itself. In the example, the opinion tuple is "<sehr enttäuscht |sn| Service>" - *'<very disappointed |sn| service>'*.

As the opinion holder in a forum in most of the cases is just the writer of a post and the insurance is mostly given directly or in the title of the forum thread, an opinion quintuple defined in [13, 20] giving the entity, the aspect, the opinion, the opinion holder and the time could be constructed in most of the cases.

If more aspects than opinion phrases are found, the opinion phrase is linked to both aspects, e.g., "Die Mitarbeiter und der Service sind sehr gut!" - *'The employees and the service are very good!'*. This results in the tuples "<sehr gut |sp| Mitarbeiter>" - *'<very good |sp| employees>'* and "<sehr gut |sp| Service>" - *'<very good |sp| service>'*.

If more opinion phrases than aspects are found, e.g., one aspect and two opinion phrases, only the nearest phrase is linked to the aspect.

4. EXPERIMENTS

4.1 Test Data

For our experiments we chose the domain 'automobile insurances'. Data from Ciao[6], which provides a review platform for a wide variety of products and services, were retrieved using a tailored crawler. To split the sentences, the Apache OpenNLP Sentence Detector was used again. The total corpus consist of about 14,000 sentences extracted from about 1,600 reviews concerning about 120 insurances. Comments to the posts were not considered. The length of the sentences used for this study was restricted to be lower than 200 characters to avoid mistakes done by the sentence tokenizer. Errors occurred for example, if a sentence delimiter is used in an improper way, e.g., "Der Service ist lausig!!!" - *'The service is lousy!!!'*, or just if the whitespace after the sentence delimiter is missing.

After these preselection steps, approximately 12,000 sentences remained.

4.2 Manual Classification of Sentences

To be able to quantify the accuracy of our algorithm, sentences were classified manually. The task to perform was to tag strong opinions expressed about aspects of insurances. The tagging was done sequentially by two persons, in the following called annotators A and B, not involved in the project and not aware of the algorithm used for the opinion mining later on. A possible result of the tagging was, e.g., "Der Service ist miserablel, aber die Mitarbeiter sind sehr nett. <Service | sn>, <Mitarbeiter | sp>" - *'The service is lousy, but the employees are very kind. <service | sn>, <employees | sp>'*.

[6]http://ciao.de/

The sentences to be tagged were preselected randomly from the test data corpus. The only selection criterion was the presence of at least one aspect per sentence. The tagging was performed until each more than 100 strong positive and strong negative expressions of opinions were found.

At the end, 221 sentences with 234 aspects remained. A strong positive attitude was tagged for 119 of them, in 115 cases the author expressed a strong negative opinion about the aspect.

As already mentioned, the tagging persons were not informed about the algorithm, especially they did not know that the opinions are searched according to opinion bearing phrases. Thus, also sentences, where the opinions were expressed in an indirect way or using idioms, were accepted. An example would be "ABC-Versicherung - vergiss es!" - *'ABC-Insurance - forget it!'*.

To quantify the agreement of the tagging persons, we calculated Cohen's kappa coefficient, getting a value of $\kappa = 0.821$ with a p-value lower than 0.001. Thus, the agreement of the tagging results was very good for the two persons doing the manual tagging. For the calculation of the accuracies, the tagging results of annotator A were taken as the reference.

5. RESULTS AND DISCUSSION

5.1 Experimental Results

For both subsets, strong positive and strong negative expressions of opinions, we calculated the accuracy. To be counted as correct, both, the detection of the tonality (strong positive or strong negative opinion uttered) and the link to the aspect, had to be correct. Using the algorithm described above, we found that 74 out of the 119 positive statements were recognized correctly, while only 17 out of the 115 negative statements could be detected in the right way.

Thus, we reach an accuracy of 62.2% for the strong positive and only an accuracy of 14.8% for the strong negative expressions of opinions.

In the following, we give examples of statements correctly detected as being positive or negative, see Section 5.1.1.

The main purpose of our work is the investigation of the cases, in which the opinion utterance can not be detected correctly. We manually analysed the error sources statement by statement. In the following Sections (5.1.2 to 5.1.10), we discuss the reasons for the failure of our algorithm. In principle, more than one error could be present for one aspect in one sentence. For example, in the sentence "Ich hasse diese bescheuerte Versicherung" - *'I hate this dumb insurance'* it could be the case that the word "bescheuert" - *'dumb'* is not in the opinion lexicon and in addition, the writer uses a verb-based phrase. However, these cases occur very rarely so statements like this are counted only once in the determination of the error sources.

Before discussing the error sources in detail, we have to say that we found no case in which a positive statement was assessed as being negative or vice versa. Such an error typically would occur, if a negation is overseen or treated in the wrong way. As we said before, we used an opinion lexicon which directly includes negation and valence shifting in the phrases, leading to the fact that there is no need for a special treatment of negation and valence shifting. We have to admit, that cases of a wrong treatment of negations can not be ruled out completely. However, these cases seem to be rare enough not to play a role in this study.

5.1.1 Statements Correctly Detected

In the following, we give some examples for statements which were recognized correctly in an ascending order of sentence complexity.

- "Toller Service!" - *'Great service!'*
- "Ich bin sehr enttäuscht von der ABC Versicherung." - *'I am very disappointed of the ABC-Insurance.'*
- "Aus meinem Bekanntenkreis, durch den ich auch letztendlich bei der ABC gelandet bin, habe ich nur Positives gehört." - *'From my friends, which brought me to the ABC, I heard only good things about it.'*
- "Sicherlich ist die ABC nicht die billigste Versicherung auf dem Markt, doch mit Abstand eine der besten, was Service, Kundendienst und Leistungen angeht." - *'For sure, the ABC is not the cheapest insurance, but by far one of the best according to the service, the customer satisfaction and the insurance benefits.'*

5.1.2 Opinions Expressed via Opinion Bearing Verbs

The opinion lexicon used just contains opinion phrases based on polar adjectives and nouns. Thus, opinions expressed using verbs can not be detected. In 19 (16%) of the positive and 43 (37.4%) of the negative statements authors used verbs to express their opinions. An example is "Insgesamt kann ich die ABC-Versicherung empfehlen" - *'All in all I recommend the ABC-insurance'*.

5.1.3 Indirect Expression of Opinions

In 4 cases (3.5%) of negative expressions, the opinion was uttered in an indirect way. An example for such an indirect statement is "Hier eine kleine Geschichte über die XYZ, meine Ex-Versicherung." - *'Here a little story about the XYZ, my ex-insurance.'*. For the sentences expressing a positive opinion, this case did not occur.

5.1.4 Opinions Uttered with Idiomatic Expressions

In 4 (3.4%) of the positive and 16 (13.9%) of the negative statements, the opinions were uttered using an idiomatic expression. An example for this is "Finger weg von der ABC Versicherung!" - *'ABC insurance - Hands off!'*.

5.1.5 Wrong Links of the Opinions to the Aspects

In our approach, we allow for more than one opinion bearing phrase and also for more than one aspect in a single sentence. Thus, the links of the phrases to the aspects can be wrong. In our sample, this error occurred in 5 cases (4.2%) of the positive and also 5 cases (4.3%) of the negative statements. An example for this is "Die Servicezeiten sind hier nicht so toll, die Kundenbetreung hingegen ist einmalig gut" - *'The service hours are not so good, the customer service, on the other hand, is brilliant'*. "nicht so toll" - *'not so good'* has a smaller distance to "Kundenbetreuung" - *'customer service'* so it is linked to the wrong aspect.

5.1.6 Phrases Missing in the Opinion Lexicon

In some cases, the opinion was expressed using an adjective or noun phrase, nevertheless, it could not be resolved as the phrase was missing in the opinion lexicon. An example for such a statement is the word "Inkompetenz" - *'incompetence'*, which was missing in the opinion list used. This error occurred in 6 (5%) of the positive and 19 (16.5%) of the negative statements.

5.1.7 Wrong Opinion Values in the Opinion Lexicon

Sometimes, the opinion phrase is included in the opinion list, but the opinion value is below the threshold for strong positive and strong negative words. This error occurred for 1 positive utterance (0.8%) and for 2 negative ones (1.7%). The associated opinion phrases were "schnell" - *'fast'* with an opinion value (OV) of 0.089, "Angst" - *'fear'* (OV $= 0.252$) and "mangelhaft" - *'insufficient'* (OV $= -0.661$).

5.1.8 Irony and Sarcasm

In 2 cases (1.7%), negative opinions were expressed using irony, an example was "Ich dachte, die XYZ ist ihr Geld wert, aber wenn es darauf ankommt, kann man sich mal wieder auf die 'Versicherung' verlassen." - *'I thought, the XYZ is worth its money, but when it comes to the crunch, again you can count on the "insurance".'*. For the positive statements, this error did not occur.

5.1.9 Spelling Mistakes and Specialities

Text sources show a wide range of the grade of correctness concerning grammar and spelling. Especially, text data retrieved from Web 2.0 platforms are often written in a very 'creative' way.
In 6 cases (5.0%) for positive and also 6 cases (5.2%) for negative statements, spelling mistakes lead to the fact that the phrases could not be recognized correctly. An example was "TOLLE Versicherung!" - *'GREAT insurance!'*. Here, the POS-tagger does not recognize "TOLLE" - *'GREAT'* as an adjective and therefore, the pattern-based phrase recognition for the application of the opinion lexicon fails.

5.1.10 Comparisions

In 4 cases (3.4%) of the positive and 1 case (0.9%) of the negative statements, the opinion values could not determined correctly due to comparisons used. An example is "Jetzt bin ich bei einer Direktversicherung, die ist um einiges günstiger als die XYZ" - *'Now I am with a direct insurance which is significantly cheaper then the XYZ'*.

5.2 Summary

We want to summarize the results of the investigation of the errors occurring during the opinion phrase extraction and the linking of the phrases to the aspects.
Our algorithm is based on an opinion lexicon including only adjective- and noun-based phrases, so up to now it is not capable of dealing with verb-based phrases. Nevertheless, we also allowed opinions expressed via verb-based phrases as we wanted to find out the fraction of the usage of verbs in expressions of opinions.
Thus, we calculate the accuracy in two ways, once including verb-based phrases (a) and once excluding them for the study (b).
Table 3 gives an overview of the frequency of the several error sources for positive statements, Table 4 for negative utterances of opinions.

We can say, that for some categories of problems a solution will be possible. It is clear, that for both, positive and negative statements, the inclusion of verb-based phrases is essential as the lack of these is the main error source in both cases.
The main error sources apart from the missing verb phrases are improper links of the phrases to the aspects due to the

Statements	Number	Percentage
Total - strong positive (a)	119	100.0%
Correctly recognized (a)	74	62.2%
Total - strong positive (b)	100	100.0%
Correctly recognized (b)	74	74.0%
Error Source	**Number**	**Percentage**
Verb-based phrases	19	16.0%
Indirect opinion expressions	0	0.0%
Idiomatic expressions	4	3.4%
Wrong links	5	4.2%
Phrases missing	6	5.0%
Wrong opinion value	1	0.8%
Irony / Sarcasm	0	0.0%
Spelling mistakes	6	5.0%
Comparisons	4	3.4%

Table 3: Statistical summary for strong positive statements, (a) including verb-based phrases and (b) excluding them.

Statements	Number	Percentage
Total - strong negative (a)	115	100.0%
Correctly recognized (a)	17	14.8%
Total - strong negative (b)	72	100.0%
Correctly recognized (b)	17	23.6%
Error Source	**Number**	**Percentage**
Verb-based phrases	43	37.4%
Indirect opinion expressions	4	3.5%
Idiomatic expressions	16	13.9%
Wrong links	5	4.3%
Phrases missing	19	16.5%
Wrong opinion values	2	1.7%
Irony / Sarcasm	2	1.7%
Spelling mistakes	6	5.2%
Comparison	1	0.9%

Table 4: Statistical summary for strong negative statements, (a) including verb-based phrases and (b) excluding them.

simple distance assignment, missing phrases in the opinion list and wrong POS-tags due to spelling mistakes. Especially for negative utterances, another main error source is the usage of idiomatic expressions.
In the following, we want to discuss possible solutions for the error sources listed above:

- Errors due to the wrong links between aspects and opinion phrases can occur in cases where two opinions are uttered about two aspects using a main and a subordinate clause, for an example see Section 5.1.5. A solution for this problem could be the use of a sentence parser to split main and sub- clauses in order to apply the distance-based linking of aspects to opinion phrases only within the splitted (sub-)clauses. We performed first experiments using the Stanford Sentence Parser[7] to split up the sentences using the German PCFG model[17, 29]. Results are looking quite promising, but it can not yet be told, in how many cases the problem of wrong links can really be solved.

- The problem of missing phrases in the opinion list could be solved to a certain extend by expanding the opinion list. Up to now, the list was constructed only using Amazon reviews. These reviews almost are written to evaluate products, only a few of them contain statements concerning services. This leads to the fact that a big part of the vocabulary used to express opinions about services is missing in the list. Thus, other review platforms, especially sources providing reviews concerning services, could be used to enrich the opinion list. For example the Ciao platform, used for this study only as a source for the statements, could be used to find additional and domain specific opinion bearing words. Moreover, for concrete applications also a manual enhancement could be feasible.

- Errors occurring due to misspelling can be one of the most serious problems when applying opinion mining on data retrieved from Web 2.0 platforms. In our study, the main effect of misspellings were wrong results in the POS-tagging step (see Section 5.1.9). The application of spell checking and correction can help to solve this problem. In some cases, were the problem is just a misuse of capital letters (see the example in Section 5.1.9), one could try to repeat the POS-tagging after replacing the capital letters by small ones. However, it is not yet clear, in how many of the cases this really solves the problem.

- Idiomatic expressions are widely used for statements expressing opinions, especially for negative utterances, see Section 5.1.4. Thus, the opinion lexicon will be extended with these idiomatic expressions.

5.3 Shortcomings and Future Work

The accuracy obtained with our approach looks a little bit sobering, especially for negative utterances of opinions. Here, we want to discuss several limitations of our work and point out the way of our future work. One main source for an improvement is the absence of verb-based phrases in the opinion lexicon used for this study. As the tonality of many verbs is highly domain-dependent, a special treatment of verbs is necessary. In our view, a special taxonomy of opinion words has to be set up for the domain under investigation to treat a high percentage of the verbs in the correct way.
At the moment, we do not resolve coreferences. This leads to the fact that sentences, in which an aspect is not directly stated but is coreferenced, e.g., using a pronoun, are not selected into our sample. An example for this would be "Die Mitarbeiter der ABC sind sehr kompetent. Sie könnten aber schneller reagieren." - *'The staff of the ABC are very competent. But they could be a little bit faster.'*. The second sentence would not pass our preselection as we only take sentences in which at least one aspect is included, see Section 4.2. Furthermore, there might be the same problem within one sentence. For example "Die Mitarbeiter sind sehr

[7] http://nlp.stanford.edu/software/lex-parser.shtml

freundlich, aber sie sind nicht kompetent." - *'The employees are very kind, but they are not competent.'* the latter opinion phrase would not be assigned to the Aspect "Mitarbeiter" - *'employees'*.

In this work, we only regard statements expressing strong opinions. We have to admit that the detection of weak positive or negative utterances is quite a lot more difficult. Thus, the accuracy drops down as soon as one includes less 'drastic' statements.

In the next steps, we are planning to include verb-based phrases into the opinion lexicon and address the problems described in Sections 5.1.4, 5.1.5 and 5.1.6, where we see the biggest chance for a significant improvement of our approach. Another task to be done is to compare our method with a machine learning approach. Furthermore, we want to apply the aspect-based opinion mining to other domains and to texts retrieved from different data sources to learn more about possible sources of problems.

6. CONCLUSIONS

In this paper, we presented our approach for the application and evaluation of aspect-based opinion mining.

We looked at strong positive and strong negative statements, written in German, about insurances, their products and services with a quite simple algorithm.

It uses a phrase-based opinion lexicon for the German language and a simple distance-based algorithm for linking the opinion phrases to the aspects. Thus, it does not require any training and is applicable to many different domains and text sources.

We showed that it is possible to reach an accuracy of 62.2% for strong positive statements, but only of 14.8% for negative ones.

The purpose of the work was the analysis of the error sources. The main shortcoming of our approach are the missing verb-based phrases in the opinion list, being responsible for about 16% of not correctly detected positive statements and for about 37% of not correctly detected negative ones.

For negative phrases, two other main error sources exist, namely missing phrases in the opinion list and the use of idiomatic expressions, which are missing in the opinion list, too.

Also other errors occur for both, positive and negative statements.

Our impression is that authors of negative statements use a wider range of possibilities for expressing their opinions, leading to the fact that the correct treatment of negative utterances of opinions is more challenging compared to the detection of positive statements.

However, we think that some of the error sources could be eliminated or at least diminished by improvements of the algorithms used.

7. ACKNOWLEDGEMENTS

The authors would like to thank all members of the Institute of Information Systems (iisys) of the University of Applied Sciences Hof for many helpful discussions and especially R. Göbel for his great efforts as to the foundation of the institute.

The project is supported by the Bavarian Ministry of Economic Affairs, Infrastructure, Transport and Technology.

The Institute of Information Systems is supported by the Foundation of Upper Franconia and by the State of Bavaria.

8. REFERENCES

[1] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 2200–2204, 2010.

[2] E. Boiy and M.-F. Moens. A machine learning approach to sentiment analysis in multilingual web texts. *Information Retrieval*, 12(5):526–558, 2009.

[3] S. Brants, S. Dipper, S. Hansen, W. Lezius, and G. Smith. The TIGER treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*, pages 24–41, 2002.

[4] J. Brooke, M. Tofiloski, and M. Taboada. Cross-Linguistic Sentiment Analysis: From English to Spanish. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 50–54, 2009.

[5] E. Cambria, C. Havasi, and A. Hussain. SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis. In *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference*, pages 202–207, 2012.

[6] E. Cambria, B. Schuller, Y. Xia, and C. Havasi. New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2):15–21, 2013.

[7] Y. Choi and C. Cardie. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 793–801, 2008.

[8] S. Clematide and M. Klenner. Evaluation and Extension of a Polarity Lexicon for German. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 7–13, 2010.

[9] X. Ding, B. Liu, and P. S. Yu. A Holistic Lexicon-Based Approach to Opinion Mining. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 231–240, 2008.

[10] A. Esuli and F. Sebastiani. Determining Term Subjectivity and Term Orientation for Opinion Mining. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 193–200, 2006.

[11] A. Esuli and F. Sebastiani. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*, pages 417–422, 2006.

[12] R. Feldman. Techniques and Applications for Sentiment Analysis. *Communications of the ACM*, 56(4):82–89, 2013.

[13] M. Hu and B. Liu. Mining and Summarizing Customer Reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, 2004.

[14] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-dependent Twitter Sentiment Classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human*

Language Technologies - Volume 1, pages 151–160, 2011.

[15] N. Jindal and B. Liu. Review Spam Detection. In *Proceedings of the 16th International World Wide Web Conference*, pages 1189–1190, 2007.

[16] N. Jindal and B. Liu. Opinion Spam and Analysis. In *Proceedings of the 1st ACM International Conference on Web Search and Data Mining*, pages 219–230, 2008.

[17] D. Klein and C. D. Manning. Accurate Unlexicalized Parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, pages 423–430, 2003.

[18] M. Klenner, S. Petrakis, and A. Fahrni. Robust Compositional Polarity Classification. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, pages 180–184, 2009.

[19] W. Lezius, R. Rapp, and M. Wettler. A Freely Available Morphological Analyzer, Disambiguator and Context Sensitive Lemmatizer for German. In *Proceedings of the 17th international conference on Computational linguistics - Volume 2*, pages 743–748, 1998.

[20] B. Liu. Sentiment Analysis and Subjectivity. In N. Indurkhya and F. Damerau, editors, *Handbook of Natural Language Processing, Second Edition*. 2010.

[21] B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *Proceedings of the 14th International World Wide Web Conference*, pages 342–351, 2005.

[22] B. Liu and L. Zhang. A Survey of Opinion Mining and Sentiment Analysis. In C. C. Aggarwal and C. Zhai, editors, *Mining Text Data*, pages 415–463. Springer US, 2012.

[23] J. Liu and S. Seneff. Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 161–169, 2009.

[24] G. A. Miller. WordNet: A Lexical Database for English. *Communications of the ACM*, 38:39–41, 1995.

[25] K. Moilanen and S. Pulman. Sentiment Composition. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, 2007.

[26] S. M. Mudambi and D. Schuff. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, 34:185–200, 2010.

[27] J.-C. Na, T. T. Thet, C. S. G. Khoo, and W. Y. M. Kyaing. Visual Sentiment Summarization of Movie Reviews. In *Proceedings of the 13th International Conference on Asia-pacific Digital Libraries*, pages 277–287, 2011.

[28] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment Classification using Machine Learning Techniques. In *Proceedings of the Conference on Empitical Methods in Natural Language Processing*, pages 79–86, 2002.

[29] A. N. Rafferty and C. D. Manning. Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines. In *Proceedings of the Workshop on Parsing German*, pages 40–46, 2008.

[30] R. Remus, U. Quasthoff, and G. Heyer. SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Ressources and Evaluation*, pages 1168–1171, 2010.

[31] S. Rill, S. Adolph, J. Drescher, D. Reinel, J. Scheidt, O. Schütz, F. Wogenstein, R. V. Zicari, and N. Korfiatis. A Phrase-Based Opinion List for the German Language. In *Proceedings of KONVENS 2012*, pages 305–313, 2012.

[32] S. Rill, J. Drescher, D. Reinel, J. Scheidt, O. Schütz, F. Wogenstein, and D. Simon. A Generic Approach to Generate Opinion Lists of Phrases for Opinion Mining Applications. In *Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM)*, 2012.

[33] A. Schiller and C. Thielen. Ein kleines und erweitertes Tagset fürs Deutsche. In *Tagungsberichte des Arbeitstreffens "Lexikon + Text", 17./18. Februar 1994, Schloß Hohentübingen*, Lexicographica Series Maior, pages 193–203. Niemeyer, Tübingen, 1995.

[34] C. Strapparava and A. Valitutti. WordNet-Affect: an Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086, 2004.

[35] H. Takamura, T. Inui, and M. Okumura. Extracting Semantic Orientations of Words using Spin Model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 133–140, 2005.

[36] T. T. Thet, J.-C. Na, and C. S. G. Khoo. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science*, 36:823–848, 2010.

[37] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, pages 173–180, 2003.

[38] P. D. Turney. 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*, pages 417–424, 2002.

[39] U. Waltinger. GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 1638–1642, 2010.

[40] W. Wei and J. A. Gulla. Sentiment Learning on Product Reviews via Sentiment Ontology Tree. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 404–413, 2010.

[41] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the Human Language Technology Conference*, pages 347–354, 2005.