SA-UZH: Verb-based Sentiment Analysis

Nora Hollenstein, Michi Amsler, Martina Bachmann, Manfred Klenner
Institute of Computational Linguistics, University of Zurich
Binzmuehlestrasse 14, CH-8050 Zurich, Switzerland
`{hollenstein,mamsler,bachmann,klenner}@ifi.uzh.ch`

Abstract

This paper describes the details of our system submitted to the SemEval-2014 shared task about aspect-based sentiment analysis on review texts. We participated in subtask 2 (prediction of the polarity of aspect terms) and 4 (prediction of the polarity of aspect categories). Our approach to determine the sentiment of aspect terms and categories is based on linguistic preprocessing, including a compositional analysis and a verb resource, task-specific feature engineering and supervised machine learning techniques. We used a Logistic Regression classifier to make predictions, which were ranked above-average in the shared task.

1 Introduction

Aspect-based sentiment analysis refers to the problem of predicting the polarity of an explicit or implicit mention of a target in a sentence or text. The SemEval-2014 shared task required sentiment analysis of laptop and restaurant reviews on sentence level and comprised four subtasks (Pontiki et al., 2014). The organizers created and shared manually labelled domain-specific training and test data sets. Two of the four subtasks dealt with determining the sentiment of a given aspect term (explicitly mentioned) or aspect category (explicitly or implicitly mentioned) in a sentence. The subtasks we participated in do not include the recognition of aspects. Given the sentence *"The sushi rolls were perfect, but overall it was too expensive."*, *"sushi rolls"* is an aspect term, and the corresponding aspect categories are *"food"* and

"price". The correct predictions would be the following:

- Subtask 2 (aspect terms): {sushi rolls → *positive*}

- Subtask 4 (aspect categories): {food → *positive*, price → *negative*}

To solve these tasks, we introduce a Logistic Regression Model for target-specific sentiment analysis. Features are derived from a fine-grained polarity lexicon, a verb resource specifying expectations and effects of the verbs functional roles, and a compositional analysis. In our experiments on the restaurant and laptop reviews data for the SemEval-2014 shared task, we found that improvements over the baseline are possible for all classes except "conflict".

2 Related Work

We focus on the question whether fine-grained linguistic sentiment analysis improves target-specific polarity classification. Existing approaches to aspect-based sentiment detection have focused on different aspects of this task, e.g. the identification of targets and their components (Popescu and Etzioni, 2005) and sentence-level composition (Moilanen and Pulman, 2007). Ding et al. (2008) and Hu and Liu (2004) produced lexicon-based approaches, which perform quite well in a large number of domains, and Blair-Goldensohn et al. (2008) combined lexicon-based methods and supervised learning. Jiang et al. (2011) used a dependency parser to generate a set of aspect dependent features for classification. For our system we built a sentiment composition resembling the one of Läubli et al. (2012), which was developed for German. Moreover, our verb resource has some similarity with the one of Neviarouskaya et al. (2009): both rely on verb classes and utilize verb-specific

behavior. However, only we specify the individual verb's (default) perspective on each role (and are, thus, able to count polar propagations). See also Reschke and Anand (2011), who describe in detail how polar (verb) complements combine to verb frame polarity (again without recording and using role perspectives as we do).

3 System Description

In this section we present the details of our sentiment analysis system. We used the same preprocessing and learning algorithm for both subtasks (2 & 4). Only the feature extraction was expanded in subtask 4 for determining the polarities of aspect categories (see section 3.3). The data sets consisted of restaurant and laptop reviews, which comprise about 3'000 manually classified target-specific sentences for each domain.

3.1 Sentiment Composition

The fundamental steps of our sentiment analysis system are parsing the sentences, rule-based sentiment analysis using a polarity lexicon and a verb resource, feature extraction and training a machine learning algorithm. In this section we will describe the composition of the lexicon as well as the structure of the sentiment composition pipeline.

Category	Example
POS strong	*"awesome"*
POS weak	*"adequate"*
NEG strong	*"catastrophe"*
NEG weak	*"demotivated"*
POS active	*"generous"*
POS passive	*"noteworthy"*
NEG active	*"rebellion"*
NEG passive	*"orphaned"*

Table 1: Additional categories in our fine-grained polarity lexicon

The same polarity lexicon was used for both domains. After mapping the polarities from the lexicon to the words and multi-word expressions, we calculated the polarity of nominal (NPs) and prepositional phrases (PPs) by means of lexical marking and the syntactic analysis of a dependency parser (Choi and Palmer, 2011). We did not implement any rules for neutral phrases, all words and phrases not marked as positive or negative are considered as neutral. In general, the polarities are propagated bottom-up to their respective heads of

the NPs/PPs in composition with the subordinates. Shifters and negation words are also taken into account. The parser output is converted into a constraint grammar (CG) format for the subsequent analysis of words and phrases. To conduct this composition of polarity for the phrases we implemented a CG with the *vislcg3* tools (VISL-group, 2013). The next stage of our sentiment detection is the verb resource, which was also implemented with the *vislcg3* tools and will be explained in the next section.

3.2 Verb-based Sentiment Analysis

In order to combine the composition of the polar phrases with verb information, we encoded the impact of the verbs on polarity using three dimensions: effects, expectations and verb polarity. While effects should be understood as the outcome instantiated through the verb, expectations can be understood as anticipated polarities induced by the verb. Effects and expectations are assigned to subjects or objects, not to the verb itself. A positive or negative verb effect propagates from the verb to a subject or object if the latter receives the polarity of the verb. For a verb expectation, the subject or object is expected to be polar and thus receives a polarity even if the sentiment composition resulted neutral (see examples below). The verb polarity as such is the evaluation of the whole verbal phrase. Moreover, we process predicative and passive verbs, adapting the effects and expectations to the syntactic structure.

Since these effects and expectations match directly to the subject and objects of a sentence, they are of great use detecting the polarity of aspect terms (which are predominantly subjects or objects). We present the following examples extracted from the training data to illustrate three dimensions annotated by the verb analysis:

- Example of a positive effect on the direct object of a sentence induced by the verb: "I love (verb_POS) *the operating system and the preloaded software* (POS_EFF)."

- Example for a negative expectation on a prepositional object induced by the verb: "[...] the guy, who constantly complains (verb_NEG) *about the noise level* (NEG_EXP)."

- Example of positive predicative effects with an auxiliary, non-polar verb: "*Ser-*

vice (POS_predicative) is (verb_PRED) great, *takeout* (POS_predicative) is (verb_PRED) good too."

Furthermore, we make a distinction between the different prepositions a verb can invoke and the succeeding semantic changes. For example, the verb *"to die"* can be annotated in three different manners, depending on the prepositional object:

1. "My phone died (verb_NEG)."

2. *"Their pizza* (POS_EFF) is to die (verb_POS) **for**."

3. "He died (verb_NEG) **of** *cancer* (NEG_EXP)."

To summarize, in addition to verb polarity, we introduce effects and expectations to verb frames, which are determined through the syntactic pattern found, the bottom-up calculated phrase polarities and the meaning of the verb itself. We manually categorized approx. 300 of the most frequent positive and negative English verbs and their respective verb frames.

Laptop reviews

Feature	Occurrences	in %
Verbs effects	367	12.05
Verb expectations	6	0.02
Predicatives	298	9.78
Polar verbs	530	17.39

Restaurant reviews

Feature	Occurrences	in %
Verbs effects	246	8.09
Verb expectations	12	0.04
Predicatives	378	12.43
Polar verbs	521	17.13

Table 2: Occurrences and percentage of sentences of annotated polar verb features in the training data of the shared task

In table 2, we illustrate the relevance of the linguistic features of this verb resource by showing in how many sentences of the training set these annotations appear. Since we merely annotated the verb frames of the most frequent English verbs, it is conceivable that this resource may have a considerably greater effect if more domain-specific verbs are modelled.

After this final sentiment composition step, all derived polarity chunks are converted into a set of features for machine learning algorithms.

3.3 Feature Extraction

In a first step of our system, the sentences are parsed, phrase polarities are calculated and verb effects and expectations are assigned. Subsequently, a feature extractor, which extracts and aggregates polar information, operates on the output. The Simple Logistic Regression classifier from *weka* Hall et al. (2009) is then trained on these features.

We developed a feature extraction pipeline that retrieves information about various polarity levels in words, syntactic functions and phrases of the sentences in the data set. In order to use our sentiment composition approach for machine learning, we extract three different sets of features, resulting in a total of 32 features for subtask 2 and 39 features for subtask 4.

In short, the feature sets are constructed as follows:

- Lexicon-based features: These features comprise simple frequency counts of positive and negative words in the sentences and binary features showing whether any positive or negative, strong or active tokens are present at all. Furthermore, these features not only include absolute counts but also token ratios.

- Composition-based features: This feature set describes the information found in nominal, prepositional and verbal phrases, such as the number of positive/negative phrase heads or predicative verb effects found. It is also possible to distinguish between features which represent frequency counts and features which represent polarity ratios.

- Target-specific features: This set includes features from the previous two sets in connection with the aspect terms, e.g. whether the aspect term has a verb expectation or whether the aspect term is the head of a negative/positive phrase, the subject or direct object, etc. In this set we also include accumulative features that represent the complete amount of polar information in connection with an aspect term.

- *(only for subtask 4)* Category-specific features: These features are based on a co-occurrence analysis of the most frequent words used in each category. That is to

say, we calculated the frequencies of all polar nouns, verbs and adjectives that appear in sentences of the same category in order to find category-specific words which have an influence on the polarity. This set includes features such as the number of category-specific words occurring in the sentence, etc.

For the classification of the aspect terms and categories of the sentences into the four classes (positive, negative, neutral and conflict), we trained a Simple Logistic Regression classifier on the features described above. We also explored other machine learning algorithms such as SVMs and artificial neural networks, however, the Logistic Regression proved to yield the best results.

4 Results & Discussion

In this section we present and discuss the results of our system in the SemEval 2014 shared task. The results of our submission for subtasks 2 and 4, compared to the majority baselines, can be found in table 3. Our system performs significantly better on restaurant reviews than on laptop reviews, probably due to the fact that our polarity lexicon comprises more restaurant-specific vocabulary than computer-specific vocabulary.

Subtask	Data	Baseline	Acc.
(2)	Laptops	47.06	58.30
(2)	Restaurants	57.8	70.98
(4)	Restaurants	59.84	73.10

Table 3: Shared-Task results for subtask 2 (aspect term polarity) and subtask 4 (aspect category polarity)

In both subtasks, calculating the polarity of the aspect terms and the aspect categories, the class *positive* scores better than the three other classes. In all data sets and all subtasks *positive* was the majority class of the four-partite classification: 42% in the aspect terms of the laptop reviews, 59% in the aspect terms and aspect categories of the laptop reviews equally (measured in the training data). Thus, it is not surprising that the most frequent error of our system is to categorize neutral aspect terms and categories as *positive*.

We do not achieve any improvements for the class *conflict*. The latter is very hard to detect, not only because this class is difficult to define but also because of the lack of training data given for this class. This could not be improved even though we included lexical features to address this particular class, for example, Boolean features showing whether an adversative conjunction is present in the sentence or whether the count of positive chunks equals the count of negative chunks in the same sentence. These features are in line with the theory that aspects are considered controversial if positive and negative occurrences are balanced and no polarity clearly prevails. Furthermore, the conflictive facet of a sentence is frequently not represented in the words (e.g. *"It has no camera, but I can always buy and install one easy."*; camera = *conflict*). Thus, it becomes challenging to generate features for this class *conflict* with a lexicon-based approach.

Furthermore, since our verb resource was newly implemented, there are still many verbs (especially domain-specific verbs) which will have to be modelled in addition to the most frequent English verbs included in the analysis by now. Another limitation of our current system is the fact that verb negation is not yet implemented: We process negation occurring in noun phrases (e.g. *"a not so tasty chicken curry"*), but not when the negation word relates to the verb (e.g. *"we didn't complain"*).

In summary, our aspect-based sentiment analysis pipeline takes into consideration many linguistic characteristics relevant for detecting opinion, and still provides the possibility to expand our compositional resources.

5 Conclusion

Given the above-average results obtained in the shared task system ranking, we conclude that the method for aspect-based sentiment analysis in review texts presented in this paper yields competitive results. We showed that the performance for this task can be improved by using linguistically motivated features for all classes except *conflict*.

We presented a supervised aspect-based sentiment analysis system to detect target-specific polarity with features derived from a fine-grained polarity lexicon, a verb resource and compositional analysis based on a dependency parser. Our results have shown that deeper linguistic analysis can positively influence the detection of target-specific polarities on sentence level in review texts.

Acknowledgements

References

Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, , and Jeff Reynar. Building a sentiment summarizer for local service reviews. In *WWW Workshop on NLP in the Information Explosion Era*, 2008.

Jinho D. Choi and Martha Palmer. Getting the most out of transition-based dependency parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, HLT '11, pages 687–692, Stroudsburg, PA, USA, 2011. ACL.

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

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.

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

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 151–160. Association for Computational Linguistics, 2011.

Samuel Läubli, Mario Schranz, Urs Christen, and Manfred Klenner. Sentiment Analysis for Media Reputation Research. In *Proceedings of KONVENS 2012 (PATHOS 2012 workshop)*, pages 274–281, Vienna, Austria, 2012.

Karo Moilanen and Stephen Pulman. Sentiment composition. In *Proceedings of RANLP-2007*, pages 378–382, Borovets, Bulgaria, 2007.

Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. *Semantically distinct verb classes involved in sentiment analysis*. IADIS AC (1), 2009.

Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, 2014.

Ana-Maria Popescu and Oren Etzioni. Extraction of product features and opinions from reviews. In *Proceedings of HLT-EMNLP-05*, pages 339–349, Vancouver, Canada, 2005.

Kevin Reschke and Pranav Anand. Extracting contextual evaluativity. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 370–374, 2011.

VISL-group. *http://beta.visl.sdu.dk/cg3.html*. Institute of Language and Communication (ISK), University of Southern Denmark, 2013.