

# UBham: Lexical Resources and Dependency Parsing for Aspect-Based Sentiment Analysis

**Viktor Pekar**

School of Computer Science  
University of Birmingham  
Birmingham, UK  
v.pekar@cs.bham.ac.uk

**Naveed Afzal**

FCIT, North Branch  
King Abdulaziz University  
Jeddah, KSA  
nafzal@kau.edu.sa

**Bernd Bohnet**

School of Computer Science  
University of Birmingham  
Birmingham, UK  
b.bohnet@cs.bham.ac.uk

## Abstract

This paper describes the system developed by the UBham team for the SemEval-2014 Aspect-Based Sentiment Analysis task (Task 4). We present an approach based on deep linguistic processing techniques and resources, and explore the parameter space of these techniques applied to the different stages in this task and examine possibilities to exploit interdependencies between them.

## 1 Introduction

Aspect-Based Sentiment Analysis (ASBA) is concerned with detection of the author's sentiment towards different issues discussed in a document, such as aspects or features of a product in a customer review. The specific ASBA scenario we address in this paper is as follows. Given a sentence from a review, identify (1) *aspect terms*, specific words or multiword expressions denoting aspects of the product; (2) *aspect categories*, categories of issues being commented on; (3) *aspect term polarity*, the polarity of the sentiment associated with each aspect term; and (4) *aspect category polarity*, the polarity associated with each aspect category found in the sentence. For example, in:

*I liked the service and the staff, but not the food.*

aspect terms are *service*, *staff* and *food*, where the first two are evaluated positively and the last one negatively; and aspect categories are SERVICE and FOOD, where the former is associated with positive sentiment and the latter with negative. It should be noted that a given sentence may contain

---

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

The research was partially supported by FP7 ICT project "Workbench for Interactive Contrastive Analysis of Patent Documentation" under grant no. FP7-SME-606163.

one, several, or no aspect terms, one, several, or no aspect categories, and may express either positive, negative, neutral, or conflicted sentiment.

While the ASBA task is usually studied in the context of documents (e.g., online reviews), peculiarities of this scenario are short input texts, complex categorization schemas, and a limited amount of annotated data. Therefore we focused on ways to exploit deep linguistic processing techniques, which we use for both creating complex classification features and rule-based processing.

## 2 Related Work

### 2.1 Aspect Term Extraction

To recognize terms that express key notions in a product or service review, a common general approach has been to extract nouns and noun phrases as potential terms and then apply a certain filtering technique to ensure only the most relevant terms remain. These techniques include statistical association tests (Yi et al., 2003), associative mining rules with additional rule-based post-processing steps (Hu and Liu, 2004), and measures of association with certain pre-defined classes of words, such as part-whole relation indicators (Popescu and Etzioni, 2005).

### 2.2 Aspect Category Recognition

Aspect category recognition is often addressed as a text classification problem, where a classifier is learned from reviews manually tagged for aspects (e.g., Snyder and Barzilay, 2007, Ganu et al., 2009). Titov and McDonald (2008) present an approach which jointly detects aspect categories and their sentiment using a classifier trained on topics discovered via Multi-Grain LDA and star ratings available in training data. Zhai et al. (2010) presented an approach based on Expectation-Maximization to group aspect expressions into user-defined aspect categories.

## 2.3 Sentence Sentiment

Lexicon-based approaches to detecting sentiment in a sentence rely on a lexicon where words and phrases are provided with sentiment labels as well as on techniques to recognize “polarity shifters”, phrases causing the polarity of a lexical item to reverse. Early work on detection of polarity shifters used surface-level patterns (Yu and Hatzivassilouglu, 2003; Hu and Liu, 2004). Moilanen and Pulman (2007) provide a logic-oriented framework to compute the polarity of grammatical structures, that is capable of dealing with phenomena such as sentiment propagation, polarity reversal, and polarity conflict. Several papers looked at different ways to use syntactic dependency information in a machine learning framework, to better account for negations and their scope (Nakagawa et al., 2010; Socher et al., 2013).

To adapt a generic sentiment lexicon to a new application domain, previous work exploited semantic relations encoded in WordNet (Kim and Hovy, 2006), unannotated data (Li et al, 2012), or queries to a search engine (Taboada et al., 2006).

## 3 Our Approach

In the following sections, we will describe our approach to each stage of the Shared Task, reporting experiments on the provided training data using a 10-fold cross-validation.

### 3.1 Aspect Term Extraction

During pre-processing training data was parsed using a dependency parser (Bohnet and Nivre, 2012), and sentiment words were recognized in it using a sentiment lexicon (see Section 6.1). Candidate terms were extracted as single nouns, noun phrases, adjectives and verbs, enforcing certain exceptions as detailed in the annotation guidelines for the Shared Task (Pontiki et al., 2014), namely:

- Sentiment words were not allowed as part of terms;
- Noun phrases with all elements capitalized and acronyms were excluded, under the assumption they refer to brands rather than product aspects;
- Nouns referring to the product class as a whole (“restaurant”, “laptop”, etc) were excluded.

Candidate terms that exactly overlapped with manually annotated terms were discarded, while those that did not were used as negative examples of aspect terms.

In order to provide the term extraction process with additional lexical knowledge, from the training data we extracted those manually annotated terms that corresponded to a single aspect category. Then the set of terms belonging to each category was augmented using WordNet: first we determined the 5 most prominent hyperonyms of these terms in the WordNet hierarchy using Resnik (1992)’s algorithm for learning a class in a semantic hierarchy that best represents selectional preferences of a verb, additionally requiring that each hypernym is at least 7 nodes away from the root, to make them sufficiently specific. Then we obtained all lexical items that belong to children synsets of these hypernyms, and further extended these lexical items with their meronyms and morphological derivatives. The resulting set of lexical items was later used as an extended aspect term lexicon. We additionally created a list of all individual lemmas of content words found in this lexicon.

For each term, we extracted the following features to be used for automatic classification:

- **Normalized form:** the surface form of the term after normalization;
- **Term lemmas:** lemmas of content words found in the term;
- **Lexicon term:** if the term is in the lexicon;
- **Lexicon lemmas ratio:** the ratio of lexicon lemmas in the term;
- **Unigram:** 3 unigrams on either side of the term;
- **Bigrams:** The two bigrams around the term;
- **Adj+term:** If an adjective depends on the term<sup>1</sup> or related to it via a link verb (“be”, “get”, “become”, etc);
- **Sentiment+term:** If a sentiment word depends on the term or related via a link verb;
- **Be+term:** If the term depends on a link verb;
- **Subject term:** If the term is a subject;

<sup>1</sup>In case the term was a multi-word expression, the relation to the head of the phrase was used.

- **Object term:** If the term is an object.

We first look at how well the manually designed patterns extracted potential terms. We are primarily interested in recall at this stage, since after that potential terms are classified into terms and non-terms with an automatic classifier. The recall on the restaurants was 70.5, and on the laptops – 56.9. These are upper limits on recall for the overall task of aspect term recognition.

Table 1 and Table 2 compare the performance of several learning algorithms on the restaurants and the laptops dataset, respectively<sup>2</sup>.

	<b>P</b>	<b>R</b>	<b>F</b>
Linear SVM	94.42	95.51	94.96
Decision Tree	94.24	92.90	93.56
Naïve Bayes	84.97	95.67	89.99
kNN ( $k=5$ )	82.71	93.50	87.76

Table 1: Learning algorithms on the aspect term extraction task, restaurants dataset.

	<b>P</b>	<b>R</b>	<b>F</b>
Linear SVM	88.14	94.07	91.00
Naïve Bayes	93.61	79.46	85.92
Decision Tree	83.87	82.99	83.39
kNN ( $k=5$ )	82.83	83.31	83.03

Table 2: Learning algorithms on the aspect term extraction task, laptops dataset.

On both datasets, linear SVMs performed best, and so they were used in the subsequent experiments on term recognition. To examine the quality of each feature used for term classification, we ran experiments where a classifier was built and tested without that feature, see Tables 3 and 4, for the restaurants and laptops datasets respectively, where a greater drop in performance compared to the entire feature set, indicates a more informative feature.

The results show the three most useful features are the same in both datasets: the occurrence of the candidate term in the constructed sentiment lexicon, the lemmas found in the term, and the normalized form of the term account.

We ran further experiments manually selecting several top-performing features, but none of the

<sup>2</sup>This and the following experiments were run on the train data supplied by the shared task organizers using 10-fold cross-validation.

	<b>P</b>	<b>R</b>	<b>F</b>
Lexicon term	91.74	95.01	93.33
Term lemmas	92.43	95.00	93.69
Normalized form	93.45	95.36	94.39
Be+term	93.99	95.28	94.63
Left bigram	94.21	95.09	94.64
All features	94.42	95.51	94.96

Table 3: Top 5 most informative features for the term extraction subtask, restaurants dataset.

	<b>P</b>	<b>R</b>	<b>F</b>
Lexicon term	88.82	88.61	88.69
Term lemmas	85.02	95.16	89.79
Normalized form	87.79	92.13	89.89
Left bigram	87.83	93.62	90.62
Term is obj	87.79	94.43	90.97
All features	88.14	94.07	91.00

Table 4: Top 5 most informative features for the term extraction subtask, laptops dataset.

configurations produced significant improvements on the use of the whole feature set.

Table 5 shows the results of evaluation of the aspect term extraction on the test data of the Shared Task (baseline algorithms were provided by the organizers). The results correspond to what can be expected based on the upper limits on recall for the pattern-based extraction of candidate terms as well as precision and recall for the classifier.

	<b>P</b>	<b>R</b>	<b>F</b>
Restaurants	77.9	61.1	68.5
Restaurants, baseline	53.9	51.4	52.6
Laptops	60.3	39.1	47.5
Laptops, baseline	40.1	38.1	39.1

Table 5: Aspect term extraction on the test data of the Shared Task.

### 3.2 Aspect Category Recognition

To recognize aspect categories in a sentence, we classified individual clauses found in it, assuming that each aspect category would be discussed in a separate clause. Features used for classification were lemmas of content words; to account for the fact that aspect terms are more indicative of aspect categories than other words, we additionally used entire terms as features, weighting them twice as much as other features. Table 6 compares the per-

formance of several learning algorithms when automatically recognized aspect terms were not used as an additional feature; Table 7 shows results when terms were used as features.

	<b>P</b>	<b>R</b>	<b>F</b>
Linear SVM	66.37	58.07	60.69
Decision Tree	58.07	51.22	53.05
Naïve Bayes	74.34	46.07	48.63
kNN ( $k=5$ )	58.65	43.77	46.57

Table 6: Learning algorithms on the aspect category recognition task, aspect terms not weighted.

	<b>P</b>	<b>R</b>	<b>F</b>
Linear SVM	67.23	59.43	61.90
Decision Tree	64.41	55.84	58.36
Naïve Bayes	78.02	49.57	52.87
kNN ( $k=5$ )	67.92	47.91	51.94

Table 7: Learning algorithms on the aspect category recognition task, aspect terms weighted.

The addition of aspect terms as separate features increased F-scores for all the learning methods, sometimes by as much as 5%. Based on these results, we used the linear SVM method for the task submission. Table 8 reports results achieved on the test data of the Shared Task.

	<b>P</b>	<b>R</b>	<b>F</b>
Restaurants	81.8	67.9	74.2
Baseline	64.8	52.5	58.0

Table 8: Aspect category extraction on the test data of the Shared Task.

### 3.3 Aspect Term Sentiment

To recognize sentiment in a sentence, we take a lexicon-based approach. The sentiment lexicon we used encodes the lemma, the part-of-speech tag, and the polarity of the sentiment word. It was built by combining three resources: lemmas from SentiWordNet (Baccianella et al., 2010), which do not belong to more than 3 synsets; the General Inquirer lexicon (Stone et al., 1966), and a subsection of the Roget thesaurus annotated for sentiment (Heng, 2004). In addition, we added sentiment expressions that are characteristic of the restaurants and laptop domains, obtained based on manual analysis of the restaurants corpus used in

(Snyder and Barzilay (2007) and the laptop reviews corpus used in (Jindal and Liu, 2008).

To detect negated sentiment, we used a list of negating phrases such as “not”, “never”, etc., and two types of patterns to determine the scope of a negation. The first type detected negations on the sentence level, checking for negative phrases at the start of the sentence; negations detected on the sentence level were propagated to the clause level. The second type of patterns detected negated sentiment within a clause, using patterns specific to the part-of-speech of the sentiment word (e.g., “AUXV + negation + VB + MAINV”, where MAINV is a sentiment verb). The output of this algorithm is the sentence split into clauses, with each clause being assigned one of four sentiment labels: “positive”, “negative”, “neutral”, “conflict”. Thus, each term was associated with the sentiment of the clause it appeared in.

On the test data of the Shared Task, the algorithm achieved the accuracy scores of 76.0 (the restaurants data, for the baseline of 64.3) and 63.6 (the laptops data, for the baseline of 51.1).

### 3.4 Category Sentiment

Recall that aspect categories were recognized in a sentence by classifying its individual clauses. Category sentiment was determined from the sentiment of the clauses where the category was found. In case more than one clause was assigned to the same category and at least one clause expressed positive sentiment and at least one – negative, such cases were classified as conflicted sentiment. This method achieved the accuracy of 72.8 (on the restaurants data), with the baseline being 65.65.

## 4 Conclusion

Our study has shown that aspect terms can be detected with a high accuracy using a domain lexicon derived from WordNet, and a set of classification features created with the help of deep linguistic processing techniques. However, the overall accuracy of aspect term recognition is greatly affected by the extraction patterns that are used to extract initial candidate terms. We also found that automatically extracted aspect terms are useful features in the aspect category recognition task. With regards to sentiment detection, our results suggest that reasonable performance can be achieved with a lexicon-based approach coupled with carefully designed rules for the detection of polarity shifts.

## References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. *SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*. Proceedings of LREC-2010.
- Bernd Bohnet and Joakim Nivre. 2012. *A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing*. Proceedings of EMNLP-CoNLL.
- Gayathree Ganu, Noémie Elhadad, and Amélié Mariani. 2009. *Beyond the Stars: Improving Rating Predictions using Review Text Content*. Proceedings of Twelfth International Workshop on the Web and Databases (WebDB 2009).
- Adrian Heng. 2004. *An exploratory study into the use of faceted classification for emotional words*. Master Thesis. Nanyang Technological University, Singapore.
- Minqing Hu and Bing Liu. 2004. *Mining opinion features in customer reviews*. Proceedings of the 9th National Conference on Artificial Intelligence (AAAI-2004).
- Nitin Jindal and Bing Liu. 2008. *Opinion Spam and Analysis* Proceedings of WWW-2008.
- Soo-Min Kim and Eduard Hovy. 2006. *Identifying and analyzing judgment opinions*. Proceedings of HLT/NAACL-2006.
- Fangtao Li, Sinno Jialin Pan, Ou Jin, Qiang Yang and Xiaoyan Zhu. 2012. *Cross-Domain Co-Extraction of Sentiment and Topic Lexicons*. Proceedings of ACL-2012.
- Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. 2010. *Dependency tree-based sentiment classification using CRFs with hidden variables*. Proceedings of NAACL/HLT-2010.
- Karo Moilanen and Stephen Pulman. 2007. *Sentiment composition*. Proceedings of the Recent Advances in Natural Language Processing (RANLP 2007).
- Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. *SemEval-2014 Task 4: Aspect Based Sentiment Analysis*. Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014).
- Ana-Maria Popescu and Oren Etzioni. 2005. *Extracting product features and opinions from reviews*. Proceedings HLT/EMNLP-2005.
- Philip Resnik. 1992. *A class-based approach to lexical discovery* Proceedings of the Proceedings of the 30th Annual Meeting of the Association for Computational Linguists.
- Benjamin Snyder and Regina Barzilay. 2007. *Multiple Aspect Ranking using the Good Grief Algorithm*. Proceedings of NAACL-2007.
- Richard Socher, Alex Perelygin, Jean Y. Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng and Christopher Potts. 2013. *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank*. Proceedings of EMNLP-2013.
- Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, MA: The MIT Press.
- Maite Taboada, Caroline Anthony, and Kimberly Voll. 2006. *Creating semantic orientation dictionaries* Proceedings of 5th International Conference on Language Resources and Evaluation (LREC).
- Ivan Titov and Ryan McDonald. 2008. *A joint model of text and aspect ratings for sentiment summarization*. Proceedings of ACL-2008.
- Liheng Xu, Kang Liu, Siwei Lai, Yubo Chen and Jun Zhao. 2013. *Mining Opinion Words and Opinion Targets in a Two-Stage Framework*. Proceedings of ACL-2013.
- Jeonghee Yi, Tetsuya Nasukawa, Razvan Bunescu, and Wayne Niblack. 2003. *Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques*. Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM-2003), pp. 423-434.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. *Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences*. Proceedings of EMNLP-03.
- Zhongwu Zhai, Bing Liu, Hua Xu and Peifa Jia. 2011. *Clustering product features for opinion mining*. Proceedings of the 4th ACM International Conference on Web Search and Data Mining, ACM, pp 347354.