

# SZTE-NLP: Aspect Level Opinion Mining Exploiting Syntactic Cues

Viktor Hangya<sup>1</sup>, Gábor Berend<sup>1</sup>, István Varga<sup>2\*</sup>, Richárd Farkas<sup>1</sup>

<sup>1</sup>University of Szeged

Department of Informatics

{hangyav,berendg,rfarkas}@inf.u-szeged.hu

<sup>2</sup>NEC Corporation, Japan

Knowledge Discovery Research Laboratories

vistvan@az.jp.nec.com

## Abstract

In this paper, we introduce our contributions to the SemEval-2014 Task 4 – *Aspect Based Sentiment Analysis* (Pontiki et al., 2014) challenge. We participated in the *aspect term polarity* subtask where the goal was to classify opinions related to a given aspect into positive, negative, neutral or conflict classes. To solve this problem, we employed supervised machine learning techniques exploiting a rich feature set. Our feature templates exploited both phrase structure and dependency parses.

## 1 Introduction

The booming volume of user-generated content and the consequent popularity growth of online review sites has led to vast amount of user reviews that are becoming increasingly difficult to grasp. There is desperate need for tools that can automatically process and organize information that might be useful for both users and commercial agents.

Such early approaches have focused on determining the overall polarity (e.g., positive, negative, neutral, conflict) or sentiment rating (e.g., star rating) of various entities (e.g., restaurants, movies, etc.) cf. (Ganu et al., 2009). While the overall polarity rating regarding a certain entity is, without question, extremely valuable, it fails to distinguish between various crucial dimensions based on which an entity can be evaluated. Evaluations targeting distinct key aspects (i.e., functionality, price, design, etc) provide important clues that may be targeted by users with different priorities concerning the entity in question, thus holding

\*The work was done while this author was working as a guest researcher at the University of Szeged

This work is licensed 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/>

much greater value in one's decision making process.

In this paper, we introduce our contribution to the SemEval-2014 Task 4 – *Aspect Based Sentiment Analysis* (Pontiki et al., 2014) challenge. We participated in the *aspect term polarity* subtask where the goal was to classify opinions which are related to a given aspect into positive, negative, neutral or conflict classes. We employed supervised machine learning techniques exploiting a rich feature set for target polarity detection, with a special emphasis on features that deal with the detection of aspect scopes. Our system achieved an accuracy of 0.752 and 0.669 for the restaurant and laptop domains, respectively.

## 2 Approach

We employed a four-class supervised (positive, negative, neutral and conflict) classifier here. As a normalization step, we converted the given texts into their lowercased forms. Bag-of-words features comprised the basic feature set for our maximum entropy classifier, which was shown to be helpful in polarity detection (Hangya and Farkas, 2013).

In the case of aspect-oriented sentiment detection, we found it important to locate text parts that refer to particular aspects. For this, we used several syntactic parsing methods and introduced parse tree based features.

### 2.1 Distance-weighted Bag-of-words Features

Initially, we used n-gram token features (unigrams and bigrams). It could be helpful to take into consideration the distance between the token in question and the mention of the target aspect. The closer a token is to an entity the more plausible that the given token is related to the aspect.

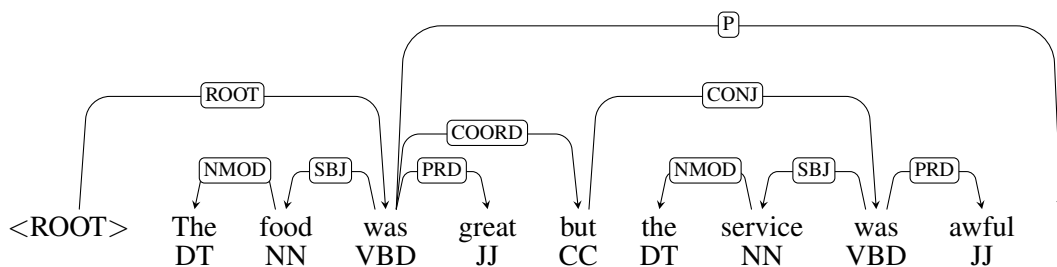


Figure 1: Dependency parse tree (MATE parser).

For this we used weighted feature vectors, and weighted each n-gram feature by its distance in tokens from the mention of the given aspect:

$$\frac{1}{e^{\frac{1}{n}|i-j|}},$$

where  $n$  is the length of the review and the values  $i, j$  are the positions of the actual word and the mentioned aspect.

## 2.2 Polarity Lexicon

To examine the polarity of the words comprising a review, we incorporated the SentiWordNet sentiment lexicon (Baccianella et al., 2010) into our feature set.

In this resource, synsets – i.e. sets of word forms sharing some common meaning – are assigned positivity, negativity and objectivity scores. These scores can be interpreted as the probabilities of seeing some representatives of the synsets in a positive, negative and neutral meaning, respectively. However, it is not unequivocal to determine automatically which particular synset a given word belongs to with respect its context. Consider the word form *great* for instance, which might have multiple, fundamentally different sentiment connotations in different contexts, e.g. in expressions such as “*great food*” and “*great crisis*”.

We determined the most likely synset a particular word form belonged to based on its contexts by selecting the synset, the members of which were the most appropriate for the lexical substitution of the target word. The extent of the appropriateness of a word being a substitute for another word was measured relying on Google’s N-Gram Corpus, using the indexing framework described in (Ceylan and Mihalcea, 2011).

We look up the frequencies of the n-grams that we derive from the context by replacing the target words with its synonyms(*great*) from various

synsets, e.g. *good* versus *big*. We count down the frequency of the phrases *food is good* and *food is big* in a huge set of in-domain documents (Ceylan and Mihalcea, 2011). Then we choose the meaning which has the highest probability, *good* in this case. This way we assign a polarity value for each word in a text and created three new features for the machine learning algorithm, which are the number of positive, negative and objective words in the given document.

## 2.3 Negation Scope Detection

Since negations are quite frequent in user reviews and have the tendency to flip polarities, we took special care of negation expressions. We collected a set of negation expressions, like *not*, *don’t*, etc. and a set of delimiters *and*, *or*, etc. It is reasonable to think that the scope of a negation starts when we detect a negation word in the sentence and it lasts until the next delimiter. If an n-gram was in a negation scope we added a *NOT* prefix to that feature.

## 2.4 Syntax-based Features

It is very important to discriminate between text fragments that are referring to the given aspect and the fragments that do not, within the same sentence. To detect the relevant text fragments, we used dependency and constituency parsers. Since adjectives are good indicators of opinion polarity, we add the ones to our feature set which are in close proximity with the given aspect term. We define proximity between an adjective and an aspect term as the length of the non-directional path between them in the dependency tree. We gather adjectives in proximity less than 6.

Another feature, which is not aspect specific but can indicate the polarity of an opinion, is the polarity of words’ modifiers. We defined a feature template for tokens whose syntactic head is present in

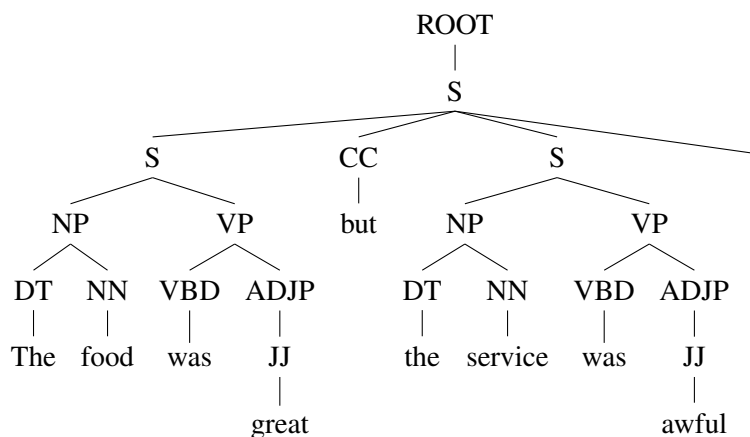


Figure 2: Constituency parse tree (Stanford parser).

our positive or negative lexicon. For dependency parsing we used the MATE parser (Bohnet, 2010) trained on the Penn Treebank (penn2malt conversion), an example can be seen on Figure 1.

Besides using words that refer to a given aspect, we tried to identify subsentences which refers to the aspect mention. In a sentence we can express our opinions about more than one aspect, so it is important not to use subsentences containing opinions about other aspects. We developed a simple rule based method for selecting the appropriate subtree from the constituent parse of the sentence in question (see Figure 2). In this method, the root of this subtree is the leaf which contains the given aspect initially. In subsequent steps the subtree containing the aspect in its yield gets expanded until the following conditions are met:

- The yield of the subtree consists of at least five tokens.
- The yield of the subtree does not contain any other aspect besides the five-token window frame relative to the aspect in question.
- The current root node of the subtree is either the non-terminal symbol PP or S.

Relying on these identified subtrees, we introduced a few more features. First, we created new n-gram features from the yield of the subtree. Next, we determined the polarity of this subtree with a method proposed by Socher et al. (2009) and used it as a feature. We also detected those words which tend to take part in sentences conveying subjectivity, using the  $\chi^2$  statistics calculated from the training data. With the help of these

words, we counted the number of opinion indicator words in the subtree as additional features. We used the Stanford constituency parser (Klein and Manning, 2003) trained on the Penn Treebank for these experiments.

## 2.5 Clustering

Aspect mentions can be classified into a few distinct topical categories, such as aspects regarding the *price*, *service* or *ambiance* of some product or service. Our hypothesis was that the distribution of the sentiment categories can differ significantly depending on the aspect categories. For instance, people might tend to share positive ideas on the price of some product rather than expressing negative, neutral or conflicting ideas towards it. In order to make use of this assumption, we automatically grouped aspect mentions based on their contexts as different aspect target words can still refer to the very same aspect category (e.g. “delicious food” and “nice dishes”).

Clustering of aspect mentions was performed by determining a vector for each aspect term based on the words co-occurring with them. 6,485 distinct lemmas were found to co-occur with any of the aspect phrases in the two databases, thus context vectors originally consisted of that many elements. Singular value decomposition was then used to project these aspect vectors into a lower dimensional ‘semantic’ space, where k-means clustering (with  $k = 10$ ) was performed over the data points. For each classification instance, we regarded the cluster ID of the particular aspect term as a nominal feature.

### 3 Results

In this section, we will report our results on the shared task database which consists of English product reviews. There are 3,000 laptop and restaurant related sentences, respectively. Aspects were annotated in these sentences, resulting in a total of 6,051 annotated aspects. In our experiments, we used maximum entropy classifier with the default parameter settings of the Java-based machine learning framework MALLET (McCallum, ).

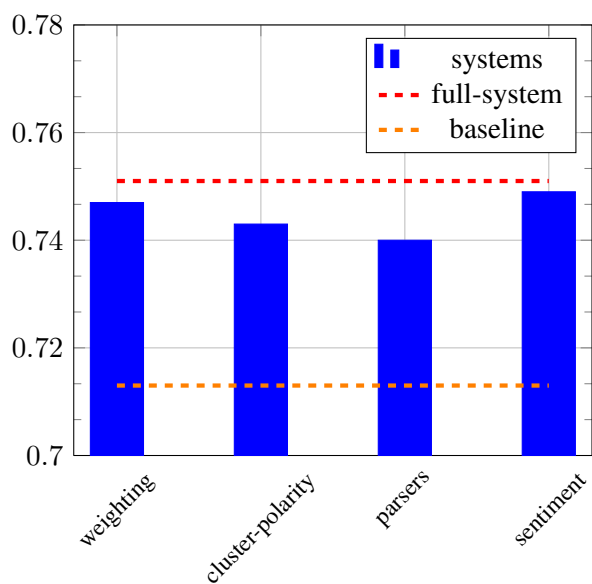


Figure 3: Accuracy on the restaurant test data.

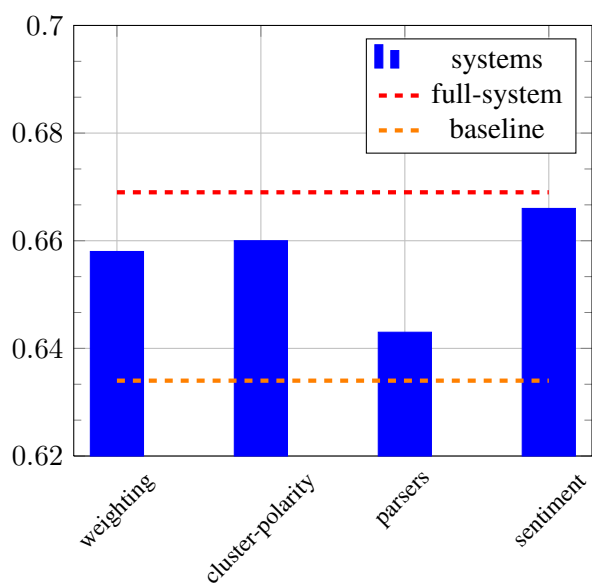


Figure 4: Accuracy on the laptop test data.

Our accuracy measured on the restaurant and laptop test databases can be seen on figures 3 and 4. On the x-axis the accuracy loss can be seen comparing to our *baseline* (n-gram features only) and *full-system*, while turning off various sets of features. Firstly, the *weighting* of n-gram features are absent, then features based on aspect clustering and words which indicate polarity in texts. Afterwards, features that are created using dependency and constituency parsing are turned off and lastly sentiment features based on the SentiWordNet lexicon are ignored. It can be seen that omitting the features based on parsing results in the most serious drop in performance. We achieved 1.1 and 2.6 error reduction on the restaurant and laptop test data using these features, respectively.

In Table 1 the results of several other participating teams can be seen on the restaurant and laptop test data. There were more than 30 submissions, from which we achieved the sixth and third best results on the restaurants and laptop domains, respectively. At the bottom of the table the official baselines for each domain can be seen.

Team	restaurant	laptop
DCU	0.809	0.704
NRC-Canada	0.801	0.704
<b>SZTE-NLP</b>	0.752	0.669
UBham	0.746	0.666
USF	0.731	0.645
ECNU	0.707	0.611
baseline	0.642	0.510

Table 1: Accuracy results of several other participants. Our system is named SZTE-NLP.

### 4 Conclusions

In this paper, we presented our contribution to the aspect term polarity subtask of the SemEval-2014 Task 4 – *Aspect Based Sentiment Analysis* challenge. We proposed a supervised machine learning technique that employs a rich feature set targeting aspect term polarity detection. Among the features designed here, the syntax-based feature group for the determination of the scopes of the aspect terms showed the highest contribution. In the end, our system was ranked as 6<sup>th</sup> and 3<sup>rd</sup>, achieving an 0.752 and 0.669 accuracies for the restaurant and laptop domains, respectively.

## Acknowledgments

Viktor Hangya and István Varga were funded in part by the European Union and the European Social Fund through the project FuturICT.hu (TÁMOP-4.2.2.C-11/1/KONV-2012-0013).

Gábor Berend and Richárd Farkas was partially funded by the "Hungarian National Excellence Program" (TÁMOP 4.2.4.A/2-11-1-2012-0001), co-financed by the European Social Fund.

## References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
- Bernd Bohnet. 2010. Top accuracy and fast dependency parsing is not a contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 89–97, Beijing, China, August. Coling 2010 Organizing Committee.
- Hakan Ceylan and Rada Mihalcea. 2011. An efficient indexer for large n-gram corpora. In *ACL (System Demonstrations)*, pages 103–108. The Association for Computer Linguistics.
- Gayatri Ganu, Noemie Elhadad, and Amelie Marian. 2009. Beyond the stars: Improving rating predictions using review text content. In *WebDB*.
- Viktor Hangya and Richard Farkas. 2013. Target-oriented opinion mining from tweets. In *Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on*, pages 251–254. IEEE.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st ACL*, pages 423–430.
- Andrew Kachites McCallum. Mallet: A machine learning for language toolkit.
- Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the International Workshop on Semantic Evaluation, SemEval '14*.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Y. Ng, and Christopher Potts. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, October.