

UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis

Tomáš Brychcín

NTIS – New Technologies
for the Information Society,
Faculty of Applied Sciences,
University of West Bohemia,
Univerzitní 8, 306 14 Plzeň
Czech Republic
brychcin@kiv.zcu.cz

Michal Konkol

NTIS – New Technologies
for the Information Society,
Faculty of Applied Sciences,
University of West Bohemia,
Univerzitní 8, 306 14 Plzeň
Czech Republic
konkol@kiv.zcu.cz

Josef Steinberger

Department of Computer
Science and Engineering,
Faculty of Applied Sciences,
University of West Bohemia,
Univerzitní 8, 306 14 Plzeň
Czech Republic
jstein@kiv.zcu.cz

Abstract

This paper describes our system participating in the aspect-based sentiment analysis task of Semeval 2014. The goal was to identify the aspects of given target entities and the sentiment expressed towards each aspect. We firstly introduce a system based on supervised machine learning, which is strictly constrained and uses the training data as the only source of information. This system is then extended by unsupervised methods for latent semantics discovery (LDA and semantic spaces) as well as the approach based on sentiment vocabularies. The evaluation was done on two domains, restaurants and laptops. We show that our approach leads to very promising results.

1 Introduction

The majority of current sentiment analysis approaches tries to detect the overall polarity of a sentence (or a document) regardless of the target entities (e.g. restaurants) and their aspects (e.g. food, price). By contrast, the ABSA (aspect based sentiment analysis) task is concerned with identifying the aspects of given target entities and estimating the sentiment polarity for each mentioned aspect.

The aspect scenario can be decomposed into two tasks: aspect extraction and aspect sentiment classification (Liu, 2012).

The task of aspect extraction is to recognize aspects of the entity and more generally can be seen as an information extraction task. The basic approach is finding frequent nouns and noun

phrases (Liu et al., 2005; Blair-Goldensohn et al., 2008; Moghaddam and Ester, 2010; Long et al., 2010). Aspect extraction can be also seen as a special case of the general information extraction problem. The most dominant methods are based on sequential learning (e.g. HMM – Hidden Markov Models (Rabiner, 2010) or CRF – Conditional Random Fields (Lafferty et al., 2001)). Another group of methods use topic models (Mei et al., 2007; Titov and McDonald, 2008; Blei et al., 2003).

Aspect sentiment classification determines whether the opinions on different aspects are positive, negative, or neutral. While lexicon-based approaches use a list of aspect-related sentiment phrases as the core resource (Ding et al., 2008; Hu and Liu, 2004), the key issue for learning methods is to determine the scope of each sentiment expression, i.e., whether it covers the aspect in the sentence (Jiang et al., 2011; Boiy and Moens, 2009).

The most of the research in aspect-level sentiment analysis has been done in English, however, there were some attempts to tackle the aspect-level task in other languages (e.g. in Czech (Steinberger et al., 2014)).

The rest of the article is organized as follows. In Section 2, we summarize the ABSA shared task (Pontiki et al., 2014). Then, we give a description of our participating system (Section 3). In Section 4, we discuss our results in the task. We participated with both the constrained and the unconstrained variants of the system.

2 The ABSA task

Datasets consisting of customer reviews with human-authored annotations identifying the mentioned aspects of the target entities and the sentiment polarity of each aspect were provided. The experiments were run in two domains: restaurant and laptop reviews.

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/>

Each team could submit two versions of systems – constrained and unconstrained. The constrained system uses only the training data and other resources (such as lexicons) for training. The unconstrained system can use additional data.

We use another definition of these types, which is not against the rules. Our constrained systems are based purely on ABSA training data, without any external knowledge such as dictionaries or rules. Our unconstrained systems use additional dictionaries, rule-based extensions and unlabeled data. From our point of view, hand-crafted dictionaries and rules are external knowledge and thus it is the same as adding external data.

The task consists of the four subtasks.

2.1 Subtask 1: Aspect term extraction

Given a set of sentences with pre-identified entities (restaurants or laptops), the task is to identify the aspect terms present in the sentence and return a list containing all the distinct aspect terms.

I liked the service and the staff, but not the food.
→ {service, staff, food}

2.2 Subtask 2: Aspect term polarity

For a given set of aspect terms within a sentence, the task is to determine the polarity of each aspect term: positive, negative, neutral or conflict (i.e., both positive and negative).

I hated their fajitas, but their salads were great.
→ {fajitas: negative, salads: positive}

2.3 Subtask 3: Aspect category detection

Given a predefined set of aspect categories, the task is to identify the aspect categories discussed in a given sentence. Aspect categories are typically coarser than the aspect terms of Subtask 1, and they do not necessarily occur as terms in the given sentence.

For example, the following categories were defined for the restaurants' domain: food, service, price, ambience and anecdotes/miscellaneous.

The restaurant was expensive, but the menu was great. → {price, food}

2.4 Subtask 4: Aspect category polarity

Given a set of pre-identified aspect categories, the task is to determine the polarity (positive, negative, neutral or conflict) of each aspect category.

The restaurant was expensive, but the menu was great. → {price: negative, food: positive}

3 System description

We use machine learning approach to all subtasks. For aspect term extraction we use CRF. For the other three tasks we use the Maximum Entropy classifier. We use the Brainy (Konkol, 2014) implementation of these algorithms.

During the data preprocessing, we use simple word tokenizer based on regular expressions. All tokens are lowercased for tasks 2 and 4.

We will firstly describe all the features used in this paper because the tasks share some of them. These features are then referenced in the descriptions of individual subtasks.

Words (W) – Word occurrence on a given position in the context window.

Bag of Words (BoW) – Occurrence of a word in a sentence (or context window).

Bigrams (B) – Bigram occurrence on a given position in the context window.

Bag of Bigrams (BoB) – Occurrence of a bigram in a sentence (or context window).

Tf-idf – Term frequency–inverse document frequency for all tokens in the sentence.

Learned Dictionary (LD) – Dictionary of terms based on training data.

Suffixes (S) – Suffix of a word (2-4 characters).

Sentiment Dictionary (SD) – Dictionary created using semi-automatic triangulation method (Steinberger et al., 2012). The score is normalized.

Senti Wordnet (SW) – See (Baccianella et al., 2010).

LDA – See Section 3.1.

Word Clusters (WC) – See Section 3.2. Cluster occurrence on a given position in the context window.

Bag of Clusters (BoC) – Same as word clusters, but without information about position.

We use two features that are not in common use in similar tasks – Latent Dirichlet Allocation and word clusters based on semantic spaces. Both these features use large amount of unlabeled data to discover latent semantics. We downloaded the restaurant reviews from <http://opentable.com>. This corpus consists of 409,665 reviews (documents) with about 27 million words. The *opentable* corpus is used as the training data for these features. Unfortunately, we did not find any large corpus for laptop domain, thus presented unsupervised features are used in restaurant domain only.

We devote the following two subsections to describe these features. Then we introduce our approach to the individual tasks.

3.1 Latent Dirichlet Allocation

The Latent Dirichlet Allocation (LDA) (Blei et al., 2003) is a topic model that is assumed to provide useful information for particular subtasks. We use LDA implementation from the MALLET (McCallum, 2002) software package. For each experiment we always train the 400 topics LDA (no significant difference was observed between different numbers of topics) with 1,000 iterations of Gibbs sampling. The hyperparameters of Dirichlet distributions were initially set to $\alpha = 50/K$, where K is the number of topics and $\beta = 0.1$. This setting is recommended by (Griffiths and Steyvers, 2004). The topic probabilities are directly used as new features to the classifier.

3.2 Word clusters

We use same approach as presented in (Brychcín and Konopík, 2014), where word clusters derived from semantic spaces improved language modeling. As recommended by these authors, we use COALS (Correlated Occurrence Analogue to Lexical Semantics) (Rohde et al., 2004) and HAL (Hyperspace Analogue to Language) (Lund and Burgess, 1996) for representing the word meaning and the Repeated Bisection algorithm for clustering. Similar approach has been already used for sentiment analysis in (Habernal and Brychcín, 2013) and (Brychcín and Habernal, 2013).

The parameters of semantic spaces are set as follows. For both semantic spaces we use a four-word context window (in both directions). HAL uses a matrix consisting of 50,000 columns, which keeps the largest amount of information. COALS uses a matrix with only 14,000 columns (as rec-

ommended by the authors of the algorithm). The SVD reduction was not used in our experiments.

Implementation of the HAL, COALS algorithms is available in an open source package S-Space (Jurgens and Stevens, 2010). For clustering, we use the implementation from the CLUTO software package (Karypis, 2003). As a measure of the similarity between two words, we use the cosine similarity of word vectors.

For both semantic spaces the word vectors are clustered into four different depths: 100, 500, 1,000, and 5,000 clusters (i.e. eight different cluster sets). The occurrences of particular clusters represent additional features to the classifiers.

3.3 Aspect term extraction

Our approach for aspect term extraction is based on Conditional Random Fields (CRF). The choice was based on similarity with the named entity recognition task, where CRF are regarded as the current state of the art (Konkol and Konopík, 2013). We use the BIO model for representing aspect terms (Ramshaw and Marcus, 1999).

The constrained feature set consists of: W , BoW , B , LD , S . It is extended by WC for the unconstrained case.

3.4 Aspect term polarity

During the detection of the aspect term polarities, the words affecting the sentiment of the aspect term are assumed to be close in most of cases. Thus we use a context window of 10 words in both directions around the target aspect term. We assume the further the word or bigram is from the target aspect term, the lower impact it has on the polarity label. To model this assumption we use a weight for each word and bigram feature taken from the Gaussian distribution according to the distance from the aspect term. The mean is set to 0 and the variance is optimized on training data.

As a feature set for the constrained approach we use only BoW , BoB and for the unconstrained approach we use BoC , SD , SW above that.

3.5 Aspect category detection

Aspect category detection is based on a set of binary Maximum Entropy classifiers, one for each class. The final decision is simply assembled from decisions of individual classifiers.

For this task we use BoW , $Tf-Idf$ for the constrained approach and add LDA , BoC for unconstrained approach.

	Team	Const.	Rank	P [%]	R [%]	F_1 [%]	Rank	ACC [%]	
Aspect terms	Restaurants	Best	1.	85.35	82.71	84.01	1.	80.95	
		UWB	U	7.	82.70	76.28	79.36	4.	77.69
		UWB	C	12.	83.28	70.28	76.23	12.	72.13
		Average	–	14-15.	76.74	67.26	70.78	18.	69.15
		Semeval Baseline	–	–	–	–	47.15	–	64.28
	Laptops	Best	–	1.	84.80	66.51	74.55	1.	70.49
		UWB	U	–	–	–	–	4.	66.67
		UWB	C	14.	77.33	49.54	60.39	10.	62.54
		Average	–	14.	68.97	50.45	56.20	16.	59.01
		Semeval Baseline	–	–	–	–	35.64	–	51.07
Aspect categories	Best	–	1.	91.04	86.24	87.58	1.	82.92	
	UWB	U	4.	84.36	78.93	81.55	8.	72.78	
	UWB	C	5.	85.09	77.37	81.04	9.	72.78	
	Average	–	11.	76.00	72.26	73.79	12-13.	69.51	
	Semeval Baseline	–	–	–	–	63.89	–	65.66	

Table 1: Comparison of our constrained (C) and unconstrained (U) system with Semeval baseline, best and average results. P , R , and F_1 denote the precision, recall and F-measure, respectively, used for measuring aspect term and category detection. ACC denotes the accuracy, used for measuring aspect term and category sentiment polarity detection.

3.6 Aspect category polarity

For this task we always take the whole sentence into account. We cannot take a limited window as we do not know where exactly the category is mentioned in the sentence. Moreover, it can be at several positions. To distinguish between different categories we again use standalone Maximum Entropy classifier for each category.

The constrained feature set consists of: *BoW*, *BoB*, *Tf-Idf*. It is extended by *BoC*, *LDA*, *SD*, *SW* for the unconstrained case.

4 Results

The ABSA task was a competition between research teams from around the world. There were 21 to 32 submitted systems for individual tasks.

We have submitted both constrained (no external knowledge, dictionaries or rules) and unconstrained systems for all tasks, except unconstrained system for aspect term extraction in the laptops domain.

Table 1 shows results of our systems (UWB) and compares them with the best and average systems as well as with the Semeval baseline. The average system is not any particular system. It is represented by average rank and metrics (metrics are averaged separately).

Our systems performed quite well. In all

tasks, we outperform the Semeval baseline system. Moreover, we are always above average (F-measure and accuracy) in all tasks. We were three times in the fourth place and our unconstrained systems were always in top ten.

Table 2 presents the 10-fold cross-validation results on restaurant training data. We can clearly see, that any of our extension (LDA, clusters, sentiment vocabularies) brings at least some improvement.

5 Conclusion

This paper covers our participation in the ABSA task of Semeval 2014. The ABSA task consists of 4 subtasks. For each subtask we propose both constrained (no external knowledge) and unconstrained approach. The constrained versions of our system are based purely on machine learning techniques. The unconstrained versions extend the constrained feature set by LDA, semantic spaces and sentiment dictionaries.

The proposed approaches achieved very good results. The constrained versions were always above average, often by a large margin. The unconstrained versions were ranked among the best systems.

	P [%]	R [%]	F_1 [%]
Constrained	68.72	82.14	74.83
Constrained + WC	76.77	82.51	79.53

(a) Aspect term extraction

	P [%]	R [%]	F_1 [%]
Constrained	74.56	80.69	77.51
Constrained + LDA	75.96	81.94	78.84
Constrained + BoC	77.01	81.42	79.16
All	77.28	81.62	79.39

(c) Aspect category extraction

	ACC [%]
Constrained	65.91
Constrained+BoC	70.05
Constrained+SD+SW	68.13
All	71.02

(b) Aspect term polarity

	ACC [%]
Constrained	66.69
Constrained+LDA	67.85
Constrained+BoC	68.61
Constrained+SD+SW	69.28
All	70.20

(d) Aspect category polarity

Table 2: 10 fold cross-validation results on the restaurants training data for individual features. P , R , and F_1 denote the precision, recall and F-measure, respectively, used for measuring aspect term and category detection. ACC denotes the accuracy, used for measuring aspect term and category sentiment polarity detection.

Acknowledgements

This work was supported by grant no. SGS-2013-029 Advanced computing and information systems, by the European Regional Development Fund (ERDF) and by project “NTIS - New Technologies for Information Society”, European Centre of Excellence, CZ.1.05/1.1.00/02.0090, and by project MediaGist, EU’s FP7 People Programme (Marie Curie Actions), no. 630786.

References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)*, Valletta, Malta, may. European Language Resources Association (ELRA).
- Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George Reis, and Jeff Reynar. 2008. Building a sentiment summarizer for local service reviews. In *Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era*.
- David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.
- Erik Boiy and Marie-Francine Moens. 2009. A machine learning approach to sentiment analysis in multilingual web texts. *Information retrieval*, 12(5):526–558.
- Tomáš Brychcín and Ivan Habernal. 2013. Un-supervised improving of sentiment analysis using global target context. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*, pages 122–128, Hissar, Bulgaria, September. INCOMA Ltd. Shoumen, BULGARIA.
- Tomáš Brychcín and Miloslav Konopík. 2014. Semantic spaces for improving language modeling. *Computer Speech & Language*, 28(1):192 – 209.
- Xiaowen Ding, Bing Liu, and Philip S. Yu. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the Conference on Web Search and Web Data Mining*.
- Thomas L. Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5228–5235, April.
- Ivan Habernal and Tomáš Brychcín. 2013. Semantic spaces for sentiment analysis. In *Text, Speech and Dialogue*, volume 8082 of *Lecture Notes in Computer Science*, pages 482–489, Berlin Heidelberg. Springer.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD ’04*, pages 168–177, New York, NY, USA. ACM.
- Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sen-

- timent classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*.
- David Jurgens and Keith Stevens. 2010. The s-space package: An open source package for word space models. In *Proceedings of the ACL 2010 System Demonstrations*.
- George Karypis. 2003. Cluto - a clustering toolkit.
- Michal Konkol and Miloslav Konopík. 2013. Crf-based czech named entity recognizer and consolidation of czech ner research. In Ivan Habernal and Václav Matoušek, editors, *Text, Speech and Dialogue*, volume 8082 of *Lecture Notes in Computer Science*, pages 153–160. Springer Berlin Heidelberg.
- Michal Konkol. 2014. Brainy: A machine learning library. In Leszek Rutkowski, Marcin Korytkowski, Rafa Scherer, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and Jacek M. Zurada, editors, *Artificial Intelligence and Soft Computing*, volume 8468 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg.
- John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of International Conference on Machine Learning*.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. In *Proceedings of International Conference on World Wide Web*.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Chong Long, Jie Zhang, and Xiaoyan Zhu. 2010. A review selection approach for accurate feature rating estimation. In *Proceedings of Coling 2010: Poster Volume*.
- Kevin Lund and Curt Burgess. 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods Instruments and Computers*, 28(2):203–208.
- Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit.
- Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of International Conference on World Wide Web*.
- Samaneh Moghaddam and Martin Ester. 2010. Opinion digger: an unsupervised opinion miner from unstructured product reviews. In *Proceeding of the ACM conference on Information and knowledge management*.
- 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 2014)*, Dublin, Ireland.
- Lawrence Rabiner. 2010. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286.
- Lance A Ramshaw and Mitchell P Marcus. 1999. Text chunking using transformation-based learning. In *Natural language processing using very large corpora*, pages 157–176. Springer.
- Douglas L. T. Rohde, Laura M. Gonnerman, and David C. Plaut. 2004. An improved method for deriving word meaning from lexical co-occurrence. *Cognitive Psychology*, 7:573–605.
- Josef Steinberger, Mohamed Ebrahim, Maud Ehrmann, Ali Hurriyetoglu, Mijail Kabadjov, Polina Lenkova, Ralf Steinberger, Hristo Tanev, Silvia Vquez, and Vanni Zavarella. 2012. Creating sentiment dictionaries via triangulation. *Decision Support Systems*, 53(4):689 – 694.
- Josef Steinberger, Tomáš Brychcín, and Michal Konkol. 2014. Aspect-level sentiment analysis in czech. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Baltimore, USA, June. Association for Computational Linguistics.
- Ivan Titov and Ryan McDonald. 2008. Modeling online reviews with multi-grain topic models. In *Proceedings of International Conference on World Wide Web*.