UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis

Tomáš Hercig[†‡]**, Tomáš Brychcín**[‡]**, Lukáš Svoboda**[†‡] **and Michal Konkol**[†‡]
[†] Department of Computer Science and Engineering, Faculty of Applied Sciences,
University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic
[‡] NTIS – New Technologies for the Information Society, Faculty of Applied Sciences,
University of West Bohemia, Technická 8, 306 14 Plzeň, Czech Republic
`tigi@kiv.zcu.cz`   `brychcin@kiv.zcu.cz`

`svobikl@kiv.zcu.cz`   `konkol@kiv.zcu.cz`

Abstract

This paper describes our system used in the Aspect Based Sentiment Analysis (ABSA) task of SemEval 2016. Our system uses Maximum Entropy classifier for the aspect category detection and for the sentiment polarity task. Conditional Random Fields (CRF) are used for opinion target extraction. We achieve state-of-the-art results in 9 experiments among the constrained systems and in 2 experiments among the unconstrained systems.

1 Introduction

The goal of Aspect Based Sentiment Analysis (ABSA) is to identify the aspects of a given target entity and estimate the sentiment polarity for each mentioned aspect. In recent years the aspect-based sentiment analysis has undergone rapid development mainly because of competitive tasks such as SemEval 2014 (Pontiki et al., 2014) and SemEval 2015 (Pontiki et al., 2015). The highest ranking participants are for SemEval 2014 (Kiritchenko et al., 2014; Brun et al., 2014; Castellucci et al., 2014; Toh and Wang, 2014; Wagner et al., 2014; Brychcín et al., 2014) and for SemEval 2015 (Saias, 2015; Toh and Su, 2015a; San Vicente et al., 2015; Zhang and Lan, 2015).

In the current ABSA task - SemEval 2016 task 5 (Pontiki et al., 2016) has attracted 29 participating teams competing in 40 different experiments among 8 languages. The task has three subtasks: Sentence-level (SB1), Text-level (SB2) and Out-of-domain ABSA (SB3). The subtasks are further divided into three slots:

- 1) Aspect Category Detection – the category consists of an entity and attribute (E#A) pair.

- 2) Opinion Target Expression (OTE)

- 3) Sentiment Polarity (positive, negative, neutral, and for SB2 conflict)

In phase A we solved slots 1 and 2. In phase B we were given the results for slots 1 and 2 and solved slot 3. We participate in 19 experiments including Chinese, English, French, and Spanish.

2 System Description

Our approach to the ABSA task is based on supervised Machine Learning. Detailed description for each experiment can be found in Section 2.2 and Section 2.3.

2.1 Features

Our system combines a large number of features to achieve competitive results. In this section we will describe the features in detail.

2.1.1 Semantics Features

We use semantics models to derive word clusters from unlabeled datasets. Similarly to (Toh and Su, 2015b) we use the `Amazon` product reviews from (Blitzer et al., 2007) and the user reviews from the `Yelp` Phoenix Academic Dataset[1] to create semantic word clusters. Additionally, we use a review `Opentable` dataset[2]. In this paper we consider the following semantics models.

---

[1] `https://www.yelp.com/dataset_challenge`
[2] downloaded from `http://opentable.com`

	Dimension	Window	Iterations
GloVe	300	10	100
CBOW	300	10	100
LDA	–	sentence	1000

Table 1: Model settings.

Global Vectors (GloVe) (Pennington et al., 2014) use the ratios of the word–word co-occurrence probabilities to encode the meanings of words.

Continuous Bag-of-Words (CBOW) (Mikolov et al., 2013) is model based on Neural Network Language Model (NNLM) that tries to predict the current word using a small context window around the word.

Latent Dirichlet Allocation (LDA) (Blei et al., 2003) discovers the hidden topics in text. We experiment with 50, 100, 200, 300, 400, and 500 topics.

The settings of the `GloVe` and `CBOW` models reflect the results of these methods in their original publications (Pennington et al., 2014; Mikolov et al., 2013). The detailed settings of all these methods are shown in Table 1.

We used the `GloVe` implementation provided on the official website[3], `CBOW` model uses the Word2Vec[4] implementation and the `LDA` implementation comes from the MALLET (McCallum, 2002) software package.

CLUTO software package (Karypis, 2003) is used for words clustering with the k-means algorithm and cosine similarity metric. All vector space models in this paper cluster the word vectors into four different numbers of clusters: 100, 500, 1000, and 5000.

The following features are based on the word clusters created using the semantic models.

Clusters (C) – The occurrence of a cluster at a given position.

Bag of Clusters (BoC) – The occurrence of a cluster in the context window.

Cluster Bigrams (CB) – The occurrence of cluster bigram at a given position.

---

[3] `http://nlp.stanford.edu/projects/glove`
[4] `https://code.google.com/p/word2vec`

Bag of Cluster Bigrams (BoCB) – The occurrence of cluster bigram in the context window.

2.1.2 Constrained Features

Affixes (A) – Affix (length 2-4 characters) of a word at a given position with a frequency > 5.

Aspect Category (AC) – extracted aspect category. We use separately the entity, attribute, and the E#A pair.

Aspect Target (AT) – listed aspect target.

Bag of Words (BoW) – The occurrence of a word in the context window.

Bag of Words filtered by POS (BoW-POS) – The occurrence of a word in the context window filtered by POS tags.

Bag of Bigrams (BoB) – The occurrence of a bigram in the context window.

Bag of Words around Verb (5V) – Bag of 5 words before verb and a bag of 5 words after verb.

Bag of 5 Words at the Beginning of Sentence (5sS) – Bag of 5 words at the beginning of a sentence.

Bag of 5 Words at the End of Sentence (5eS) – Bag of 5 words at the end of a sentence.

Bag of Head Words (BoHW) – bag of extracted head words from the sentence parse tree.

Emoticons (E) We used a list of positive and negative emoticons (Montejo-Ráez et al., 2012). The feature captures the presence of an emoticon within the text.

Head Word (HW) – extracted head word from the sentence parse tree.

Character N-gram (ChN) – The occurrence of character n-gram at a given position.

Learned Target Dictionary (LTD) – presence of a word from learned[5] dictionary of aspect terms.

Learned Target Dictionary by Category (LTD-C) – presence of a word from the learned dictionary[5] of aspect terms grouped by category.

---

[5] from training data

N-gram (N) – The occurrence of n-gram in the context window.

N-gram Shape (NSh) – The occurrence of word shape n-gram in the context window. We consider unigrams with frequency >5 and bigrams, trigrams with frequency > 20.

Paragraph Vectors (P2Vec) is an unsupervised method of learning text representation (Le and Mikolov, 2014). Resulting feature vector has a fixed dimension while the input text can be of any length. The model is trained on the *One billion word benchmark* presented in (Chelba et al., 2013), resulting vectors[6] are used as features for a sentence. We use the implementation by Řehůřek and Sojka (2010).

POS N-gram (POS-N) – The occurrence of POS n-gram in the context window.

Punctuation (P) – The occurrence of a question mark, an exclamation mark or at least two dots in the context window.

Skip-bigram (SkB) – Instead of using sequences of adjacent words (n-grams) we used skip-grams (Guthrie et al., 2006; Reyes et al., 2013), which skip over arbitrary gaps. We consider skip-bigrams with 2 to 5 word skips and remove skip-grams with a frequency ≤ 5.

Target Bag of Words (T-BoW) – BoW containing parent, siblings, and children of the target from the sentence parse tree.

TF-IDF (TF-IDF) – Term frequency - inverse document frequency of a word computed from the training data.

Verb Bag of Tags (V-BoT) – Bag of syntactic dependency tags of parent, siblings, and children of the verb from the sentence parse tree.

Verb Bag of Words (V-BoW) – Bag of words for parent, siblings, and children of the verb from the sentence parse tree.

---

[6]Vector dimension has been set to 300.

Word Shape (WSh) – we assign words into one of 24 classes[7] similar to the function specified in (Bikel et al., 1997).

Words (W) – The occurrence of word at a given position (e.g. previous word).

2.1.3 Unconstrained Features

Dictionary (DL) – presence of a word from dictionary extracted from the Annotation Guidelines for Laptops.

Dictionary (DR) – presence of a word from dictionary extracted from the Annotation Guidelines for Restaurants.

Enchanted Dictionary (ED) – presence of a word from a dictionary extracted from website[8].

Group of Words from ED (EDG) – presence of any word from a group from the ED dictionary.

Dictionary of Negative Words (ND) – presence of any negative word from the negative words list[9].

Sentiment (S) – this is a union of features dealing with sentiment. It consists of *BoG* features where the groups correspond to various sentiment lexicons. We used the following lexicon resources: Affinity lexicon (Nielsen, 2011), Senticon (Cruz et al., 2014), dictionaries from (Steinberger et al., 2012), MICRO-WNOP (Cerini et al., 2007), and the list of positive or negative opinion words from (Liu et al., 2005). Additional feature includes the output of Stanford CoreNLP (Manning et al., 2014) v3.6 sentiment analysis package by (Socher et al., 2013).

2.2 Phase A

Sentence-level Category (SB1, slot 1) We use maximum entropy classifier for all classes. Then a threshold t is used to decide which categories will be assigned by the classifier.

---

[7]We use edu.stanford.nlp.process.WordShapeClassifier with the WORDSHAPECHRIS1 setting.

[8]http://www.enchantedlearning.com/wordlist/

[9]http://dreference.blogspot.cz/2010/05/negative-ve-words-adjectives-list-for.html

Chinese We used identical features for both domains (*BoB, BoHW, BoW, ChN, N*), where *ChN* ranges from unigram to 4-gram and *ChN* with frequency < 20 are removed and *N* ranges from unigram to trigram and *ChN* with frequency < 10 are removed. The threshold was set to $t = 0.1$.

Spanish For Spanish we used the following features: *5V, 5eS, BoB, BoHW, BoW, BoW-POS, ChN, V-BoT*, where *5V* considers only adjective, adverb, and noun, *5eS* considers adjectives and adverbs with frequency > 5, *ChN* ranges from unigram to 4-gram and *ChN* with frequency < 20 are removed, *BoW-POS* is used separately for adverbs, nouns, verbs, and a union of adjectives, adverbs, nouns, and verbs, *V-BoT* is used separately for adverbs, nouns, and a union of adjectives and adverbs while reducing feature space by 50 occurrences. The threshold was set to $t = 0.2$.

English English features employ lemmatization. The threshold was set to $t = 0.14$. Common features for all experiments in this task are *5V, 5eS, BoB, BoHW, BoW, BoW-POS, P, TF-IDF, V-BoT*, where *5V* considers only adjective, adverb, and noun, *5eS* filters only adjective and adverb, *BoW-POS* contains adjectives, adverbs, nouns, and verbs, *V-BoT* filters adjectives and adverbs with frequency > 20.

The unconstrained model for the Laptops domain additionally uses *BoC, BoCB, DL, ED, P2Vec BoC* and *BoCB* include the `GloVe` and `CBOW` models computed on the `Amazon` dataset.

The constrained model for the restaurant domain additionally uses *5sS, ChN, LTD, LTD-C, P2Vec 5sS* filters only adjective and adverb, *ChN* in this case means character unigrams with frequency > 5. This model also considers separate *BoW-POS* features for groups for adverbs, nouns and verbs.

The unconstrained model for the restaurant domain uses *BoC, BoCB, DR, LDA, ND, NSh* on top of the previously listed features for the constrained model.

BoC and *BoCB* include the `GloVe`, `CBOW`, and `LDA` models computed on the `Yelp` dataset and `CBOW` model computed on the `Opentable` dataset.

Sentence-level Target (SB1, slot 2) Similarly to (Brychcín et al., 2014), we have decided to use Conditional Random Fields (CRF) (Lafferty et al., 2001)

to solve this subtask. The context for this task is defined as a five word window centred at the currently processed word. English features for this subtask employ lemmatization.

The baseline feature set consists of *A, BoB, BoW-POS, HW, LTD, LTD-C, N, POS-N, V-BoT, W, WSh*. *BoW-POS* contains adjectives, adverbs, nouns, verbs, and a union of adverbs and nouns. We consider *POS-N* with frequency > 10. *V-BoT* includes adverbs, nouns, and a union of adjectives, adverbs, nouns, and verbs.

In the unconstrained model, we extend this with the semantic features *C, CB* (created using the `CBOW` model computed on the `opentable` dataset) and with lexicons *DR, EDG*.

Sentence level Category & Target (SB1, slot 1&2) We firstly assign targets as described above and then combine them with the best five candidates for aspect category[10]. We also add aspect categories without target. This produces too many combinations thus we need to filter the unlikely opinions. We remove the opinions without target in a sentence where the aspect category is already present with a target. When there is only one target and one aspect category in a sentence we combine them into a single opinion.

Text-level Category (SB2, slot 1) We used the baseline algorithm: the predicted sentence-level tuples (SB1, slot 1) are copied to text-level and duplicates are removed.

2.3 Phase B

Sentence level Sentiment Polarity (SB1, slot 3) Our sentiment polarity detection is based on the Maximum Entropy classifier, which works very well in many NLP tasks, including document-level sentiment analysis (Habernal et al., 2014).

Chinese We used identical features for both domains (*5V, 5eS, 5sS, AC, BoB, BoHW, BoW, BoW-POS, ChN, N, NSh, P, SkB, V-BoT, V-BoW*), where *5V* considers adjectives and adverbs with frequency > 5, *5eS* and *5sS* contain adjectives, adverbs, nouns, and verbs, *BoW-POS* is used separately for adjectives and adverbs, *ChN* ranges from unigram to 5-

---

[10]We use the same settings and approach as in the sentence-level category detection (SB1 slot 1).

gram and *ChN* with frequency < 5 are removed, *N* ranges from unigram to 5-gram and *ChN* with frequency < 2 are removed, *V-BoT* is used separately for verbs, and a union of adjectives and adverbs, *V-BoW* is used separately for adjectives, adverbs, verbs, a union of adjectives and adverbs and a union of adjectives, adverbs, nouns, and verbs, while reducing feature space by 2 occurrences.

French We employ lemmatization for French. The first constrained model includes the following features: *AC, BoB, BoHW, BoW, BoW-POS, ChN, LTD, LTD-C, N, NSh, P, SkB, V-BoT*, where *BoW-POS* is used separately for adjectives, adverbs and a union of adjectives, adverbs, nouns, and verbs, *ChN* ranges from unigram to 5-gram and *ChN* with frequency ≤ 5 are removed, *N* ranges from unigram to 5-gram and *N* with frequency < 2 are removed, *V-BoT* is used separately for verbs, and a union of adjectives and adverbs.

The second constrained model additionally uses *5V, 5eS, 5sS, AT, T-Bow, V-BoW*, where *5V* considers only adjective and adverb, *5eS, 5sS* considers adjective, adverb, noun, and verb, *T-BoW* is used for adjectives, adverbs, nouns, and verbs, *V-BoW* is used separately for adjectives, adverbs, verbs, a union of adjectives and adverbs and a union of adjectives, adverbs, nouns, and verbs, while reducing feature space by 2 occurrences.

Spanish We employ lemmatization for Spanish. We used the following features: *5V, 5eS, AC, BoB, BoHW, BoW, BoW-POS, E, ChN, LTD, LTD-C, N, NSh, P, SkB, T-Bow, V-BoT, V-BoW*, where *5V* considers only adjective and adverb, *5eS* considers adjective, adverb, noun, and verb, *BoW-POS* is used separately for adjectives, adverbs, *ChN* ranges from unigram to 5-gram and *ChN* with frequency ≤ 5 are removed, *N* ranges from unigram to 5-gram and *N* with frequency < 2 are removed, *T-BoW* is used for adjectives, adverbs, nouns, and verbs, *V-BoT* is used separately for verbs, and a union of adjectives and adverbs, *V-BoW* is used separately for adjectives, adverbs, verbs, a union of adjectives and adverbs and a union of adjectives, adverbs, nouns, and verbs, while reducing feature space by 2 occurrences.

English We use lemmatization in this subtask. Common features for all experiments in this task

are *5V, 5eS, 5sS, AC, AT, BoB, BoHW, BoW, BoW-POS, E, ChN, LTD, LTD-C, N, NSh, P, SkB, V-BoT, V-BoW*, where *5V* considers adjectives and adverbs, *5eS, 5sS* consists of adjectives, adverbs, nouns, and verbs, , *BoW-POS* contains adjectives and adverbs, *N* ranges from unigram to 5-gram and *N* with frequency < 2 are removed, *V-BoT* is used separately for verbs, and a union of adjectives and adverbs, *V-BoW* is used separately for adjectives, adverbs, verbs, a union of adjectives and adverbs and a union of adjectives, adverbs, nouns, and verbs, while reducing feature space by 2 occurrences.

The unconstrained model for the Laptops domain additionally uses *BoC, BoCB, ED, S BoC* and *BoCB* include the `GloVe` and `CBOW` models computed on the `Amazon` dataset.

The constrained model for the restaurant domain additionally uses *T-Bow, TF-IDF*, where *T-BoW* is used for adjectives, adverbs, nouns, and verbs.

The unconstrained model for the restaurant domain uses *BoC, BoCB, ED, ND, S* on top of the previously listed features for the constrained model.

BoC and *BoCB* include the `GloVe` and `CBOW` models computed on the `Yelp` dataset and `CBOW` model computed on the `Opentable` dataset.

Text-level Sentiment Polarity (SB2, slot 3) The baseline algorithm traverses the predicted sentence-level tuples of the same category and counts the respective polarity labels (positive, negative or neutral). Finally, the polarity label with the highest frequency is assigned to the text-level category. If there are not any sentence-level tuples of the same category the polarity label is determined based on all tuples regardless of the category.

Our improved algorithm contains an additional step, that assigns polarity for cases (categories) with more than one sentence-level polarity labels. The resulting polarity is determined by the following algorithm:

```
if(catPolarity == lastPolarity){
  assign lastPolarity;
}else if(catPolarity == entPolarity){
  assign entPolarity;
}else{
  assign CONFLICT;
}
```

where `catPolarity` is the polarity label with the highest frequency for the given category

| Domain | Lang | Subtask | Constrained | | | | Unconstrained | | | |
| | | | Category | | Sentiment | | Category | | Sentiment | |
			Rank	$F_1[\%]$	Rank	$ACC[\%]$	Rank	$F_1[\%]$	Rank	$ACC[\%]$
Restaurants	EN	SB1	3.	67.8	2.	81.8	8.	68.2	9.	81.7
Laptops	EN	SB1	1.	47.9	3.	73.8	7.	47.3	10.	73.8
Restaurants	EN	SB2	1.	81.0	1.	80.9	3.	80.2	1.	81.9
Laptops	EN	SB2	1.	60.5	1.	74.5	2.	59.7	1. - 2.	75.0
Restaurants	FR	SB1	–	–	2.	75.3	–	–	–	–
Cameras	CH	SB1	1.	36.3	3.	77.8	–	–	–	–
Phones	CH	SB1	1.	22.5	3.	72.0	–	–	–	–
Restaurants	SP	SB1	3.	62.0	2.	81.3	–	–	–	–
Restaurants	SP	SB2	3.	73.7	1.	77.2	–	–	–	–

| Domain | Lang | Subtask | Target | | Category & Target | | Target | | Category & Target | |
			Rank	$F_1[\%]$	Rank	$F_1[\%]$	Rank	$F_1[\%]$	Rank	$F_1[\%]$
Restaurants	EN	SB1	1.	66.9	4.	41.1	3.	67.1	6.	41.1

Table 2: Achieved ranks and results (in %) by UWB for all submitted systems.

(E#A touple), `entPolarity` is the polarity label with the highest frequency for the entity E and `lastPolarity` is the last seen polarity label for the given category. This follows our believe that the last polarity tends to reflect the final sentiment (opinion) toward the aspect category.

2.4 System Settings

For all experiments we use Brainy (Konkol, 2014) machine learning library.

Data preprocessing includes lower-casing and in some cases lemmatization.

We utilize parse trees, lemmatization and POS tags from the Stanford CoreNLP (Manning et al., 2014) v3.6 framework. We chose it because it has support for Chinese, English, French, and Spanish.

3 Results and Discussion

As shown in the Table 2 we achieved very satisfactory results especially for the constrained experiments.

In the English sentence level Laptops domain our constrained method was slightly better than the unconstrained one (by 0.6%). We believe this is not a significant deviation.

The baseline algorithm for text-level category (SB2, slot 1)[11] achieves an F_1 score of 96.08% on the Laptops domain and 97.07% on the Restaurants domain for English. When we add the corresponding general class for the

given domain (e.g. `RESTAURANT#GENERAL` and `LAPTOP#GENERAL`) the algorithm achieves an F_1 score of 96.87% on the Laptops domain and 99.75% on the Restaurants domain for English.

The baseline algorithm for text-level sentiment polarity (SB2, slot 3)[11] achieves an Accuracy of 86.8% on the Laptops domain and 89.6% on the Restaurants domain for English, while the improved algorithm achieves an Accuracy of 94.5% on the Laptops domain and 97.3% on the Restaurants domain for English.

4 Conclusion

We competed in 19 constrained experiments and won 9 of them. In the other 10 cases we have reached at worst the 4th place. Our unconstrained systems participated in 10 experiments and achieved 5 ranks ranging from the 1st to 3rd place.

In the future we aim to explore the benefits of using neural networks for sentiment analysis.

Acknowledgements

[11]Using the sentence-level gold test data.

References

Daniel M Bikel, Scott Miller, Richard Schwartz, and Ralph Weischedel. 1997. Nymble: a high-performance learning name-finder. In *Proceedings of the fifth conference on Applied natural language processing*, pages 194–201. Association for Computational Linguistics.

D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:2003.

John Blitzer, Mark Dredze, Fernando Pereira, et al. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447.

Caroline Brun, Diana Nicoleta Popa, and Claude Roux. 2014. XRCE: Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 838–842, Dublin, Ireland, August. Association for Computational Linguistics.

Tomáš Brychcín, Michal Konkol, and Josef Steinberger. 2014. UWB: Machine learning approach to aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 817–822. Association for Computational Linguistics.

Giuseppe Castellucci, Simone Filice, Danilo Croce, and Roberto Basili. 2014. UNITOR: Aspect based sentiment analysis with structured learning. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 761–767, Dublin, Ireland, August. Association for Computational Linguistics.

Sabrina Cerini, Valentina Compagnoni, Alice Demontis, Maicol Formentelli, and G Gandini. 2007. Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. *Language resources and linguistic theory: Typology, second language acquisition, English linguistics*, pages 200–210.

Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge, Thorsten Brants, Phillipp Koehn, and Tony Robinson. 2013. One billion word benchmark for measuring progress in statistical language modeling. *arXiv preprint arXiv:1312.3005*.

Fermín L Cruz, José A Troyano, Beatriz Pontes, and F Javier Ortega. 2014. Building layered, multilingual sentiment lexicons at synset and lemma levels. *Expert Systems with Applications*, 41(13):5984–5994.

David Guthrie, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks. 2006. A closer look at skip-gram modelling. In *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)*, pages 1–4.

Ivan Habernal, Tomáš Ptáček, and Josef Steinberger. 2014. Supervised sentiment analysis in Czech social media. *Information Processing & Management*, 50(5):693–707.

G. Karypis. 2003. CLUTO—A clustering toolkit.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland, August. Association for Computational Linguistics.

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-Verlag, Berlin.

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*.

Quoc V. Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 1188–1196.

Bing Liu, Minqing Hu, and Junsheng Cheng. 2005. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pages 342–351. ACM.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Andrew Kachites McCallum. 2002. MALLET: A machine learning for language toolkit.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

A. Montejo-Ráez, E. Martínez-Cámara, M. T. Martín-Valdivia, and L. A. Ureña López. 2012. Random walk weighting over sentiwordnet for sentiment polarity detection on twitter. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, pages 3–10, Stroudsburg, PA, USA. Association for Computational Linguistics.

Finn Årup Nielsen. 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. In

Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages. 93-98.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Association for Computational Linguistics, Denver, Colorado*, pages 486–495.

Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval '16, San Diego, California, June. Association for Computational Linguistics.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA. `http://is.muni.cz/publication/884893/en`.

Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, 47(1):239–268.

José Saias. 2015. Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12. Association for Computational Linguistics.

Inaki San Vicente, Xabier Saralegi, Rodrigo Agerri, and Donostia-San Sebastián. 2015. Elixa: A modular and flexible absa platform. *SemEval-2015*, page 748.

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.

In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642. Citeseer.

Josef Steinberger, Mohamed Ebrahim, Maud Ehrmann, Ali Hurriyetoglu, Mijail Kabadjov, Polina Lenkova, Ralf Steinberger, Hristo Tanev, Silvia Vzquez, and Vanni Zavarella. 2012. Creating sentiment dictionaries via triangulation. *Decision Support Systems*, 53(4):689 – 694.

Zhiqiang Toh and Jian Su. 2015a. Nlangp: Supervised machine learning system for aspect category classification and opinion target extraction.

Zhiqiang Toh and Jian Su. 2015b. NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 496–501, Denver, Colorado, June. Association for Computational Linguistics.

Zhiqiang Toh and Wenting Wang. 2014. DLIREC: Aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 235–240, Dublin, Ireland, August. Association for Computational Linguistics.

Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. 2014. DCU: Aspect-based polarity classification for SemEval Task 4. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 223–229, Dublin, Ireland, August. Association for Computational Linguistics.

Zhihua Zhang and Man Lan. 2015. Ecnu: Extracting effective features from multiple sequential sentences for target-dependent sentiment analysis in reviews. *SemEval-2015*, page 736.