Aspect Based Sentiment Analysis- A Survey

Abhishek Sethi[a], **Pushpak Bhattacharyya**[a,b]
[a]Indian Institute of Technology Bombay, India
[b]Indian Institute of Technology Patna, India
{abhisheksethi505,pushpakbh}@gmail.com

Abstract

Analyzing user opinions has always been an integral part of information processing and thus Sentiment Analysis has been a very active research area since the last decade. Sentiment Analysis is done at different levels of granularity ranging from coarse grained Document level to fine grained Aspect Level. In this paper we describe the most granular form of analyzing opinions i.e Aspect Based Sentiment Analysis. In particular we focus on shared task for ABSA at International Workshop on Semantic Evaluation (SemEval- 2016) (Pontiki et al., 2016).

1 Introduction

Aspect Based Sentiment Analysis is fine grained sentiment analysis. A sentence may contain multiple opinions about different entities and we need to find each of them. We now give the quintuple definition of ABSA.

$$S = (e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

where:
e_i is the entity under consideration
a_{ij} is a particular aspect or feature of the entity
h_k is an sentiment holder
t_l is the time at which the sentiment is held
s_{ijkl} is the sentiment about aspect a_{ij} of entity e_i held by h_k at time t_l.
For example, the sentence, ”The cuisines were excellent but the service was slow.” is a review about a restaurant. It conveys sentiment about different aspects of restaurant rather than restaurant as a whole. The sentence is positive about the ”food” and negative about the ”service”. So, the tuples will be:

- (”restaurant”, ”food”, ”positive”, ”review_id”, ”review_date”)
- (”restaurant”, ”service”, ”negative”, ”review_id”, ”review_date”)

The objective is to find all these quintuples. ABSA was introduced as a shared task for the first time in SemEval- 2014 (SE-ABSA14) (Pontiki et al., 2014). In SemEval- 2014, English reviews annotated at the sentence level with aspect terms (e.g., mouse, pizza) and their polarity for the laptop and restaurant domains, as well as coarser aspect categories (e.g., food) and their polarity only for restaurants was used. The ABSA shared task at SemEval- 2015 (Pontiki et al., 2015) built upon SE-ABSA14 and consolidated its subtasks into a unified framework in which all the identified constituents of the expressed opinions (i.e., aspects, opinion target expressions and sentiment polarities) meet a set of guidelines and are linked to each other within sentence-level tuples (Pontiki et al., 2015). These tuples are important since they indicate the part of text within which a specific opinion is expressed. However, a user might also be interested in the overall rating of the text towards a particular aspect. Therefore, in addition to sentence-level annotations, SE-ABSA16 accommodated also text-level ABSA annotations and provided the respective training and testing data. There were tasks for multiple domains including restaurants, laptops, phones etc and different languages including English, French, Dutch etc.

2 Task Description

In this section, we present the ABSA shared task at SemEval 2016. There were multiple subtasks and slots in the task which are as follows:

- Subtask 1 (SB1): Sentence-level ABSA. In this task, we need to identify the following pieces of information from the review:
  - Slot1: Aspect Category
    System should identify entity E and attribute A pair towards which an opinion is expressed in the given review. E and A are chosen from predefined set of entity types (e.g., FOOD, DRINKS) and attribute labels (e.g., PRICE, QUALITY). The E, A sets for each domain are provided. For example, The soup is very tasty. belongs to {FOOD#QUALITY}
  - Slot2: Opinion Target Expression (OTE). Systems need to identify the exact target (OTE) of opinion in the sentence which is used to refer to the E#A pair . The OTE is defined by its starting and ending offsets. When there is no explicit mention of the entity, the slot takes the value null. For e.g, in the sentence, The fruit cake was very tasty., fruit cake is the target of opinion.
  - Slot3: Sentiment Polarity. Input will be a E#A and a opinion target pair. Each of these pairs should be assigned polarity between positive, negative, neutral. An example of opinion tuple with Slot 1- 3 values from the restaurants domain is as follows : The soup was very tasty and music was awesome. { category= FOOD# QUALITY , target= soup, from: 4, to: 8, polarity= positive}, {category= AMBIENCE 4,target = music, from: 28, to: 33, polarity= positive}
- Subtask 2 (SB2): Text-level ABSA
  In this sub-task, from a review(multiple sentences)- the goal is to identify a set of cat, pol tuples that summarize the opinions expressed in the review. cat can be assigned the same values as in SB1 (E#A tuple), while pol can be set to positive, negative, neutral, or conflict. For example, for the review text The So called laptop Runs to Slow and I hate it! Do not buy it! It is the worst laptop ever , a system should return the following opinion tuples: {cat: laptop#general, pol: negative}, {cat: laptop#operation_performance, pol: negative}
- Out-of domain ABSA
  In SB3, no training data was provided and teams could test their systems on an unseen domain.

3 Phases & Baseline Measures

The task consisted of two phases. In the first phase (Phase A), slot 1 and slot 2 subtasks were evaluated. For SB2 the respective text-level categories had to be identified. In the second phase (Phase B), the gold annotations for the test sets of Phase A were provided and participants had to return the respective sentiment polarity values (Slot3). We describe below the details of data, evaluation criterion and baselines for each of the slots.

- Slot 1- Aspect Category
  The E, A sets for each domain are provided. For example, the sentence, The soup is very tasty. belongs to {FOOD#QUALITY}. Evaluation measure is Micro-F1. The duplicate occurrences of categories were ignored. The baseline model used Support Vector Machine (SVM) with linear kernel. 1,000 most frequent unigrams of the training data excluding stopwords. The category value (e.g., service#general) of the tuple is used as the correct label of the feature vector. Similarly, for each test sentence s, a feature vector is built and the trained SVM is used to predict the probabilities of assigning each possible category to s (e.g., service#general, 0.2, restaurant#general, 0.4. Then, a threshold is used to decide which of the categories will be assigned to s.

- Slot 2- Opinion Target Extraction
  In the baseline model, for each category a list of OTEs is created. For e.g, for category ”FOOD#QUALITY” , the list of OTEs might be ”pizza”, ”chicken roll” etc. Then, for a test review s and an assigned category c from slot 1, we find in s the first occurrence of each OTE of cs list. The first OTE from c’s list to be found in s is considered as target for that sentence. If no target occurrences are found, the slot is assigned the value null. The calcu-

lation for each sentence considered only distinct targets and discarded null targets. The evaluation measure used is Micro-F1.

- Slot 3- Aspect Sentiment Polarity
In this subtask, SVM with linear kernel is used. Again, as in Slot1, ngram features are used. To incorporate the category information, an integer feature indicating the category information is used. The correct label for the extracted training feature vector is the corresponding polarity value (e.g., positive). Then, for each tuple category, OTE of a test sentence s, a feature vector is built and classified using the trained SVM. The evaluation measure is the accuracy of each system, defined as the number of correctly predicted polarity labels of the (gold) aspect categories, divided by the total number of the gold aspect categories.

4 Approaches to Different Subtasks In SemEval- 2016

4.1 Aspect Categorization

This is a multi-label classification problem. A review can contain sentiment about multiple categories.

- NLangp (Toh and Su, 2016) uses feed forward neural network to tackle this problem. They train feed forward neural network for for each entity attribute pair i.e a total of 12 classifiers for Restaurant domain. A threshold is set for each classifier and output of classifier in added into the set of categories of that sentence if the output probability of that classifies is larger than the threshold. The set of features used were bigrams, current word, name lists(restaurant domain)(for membership testing) and word clusters(clarke and brown)(Clark, 2000). They also train CNN to enhance the previous features usnig CNN probabilites.

- Another technique in NileTMRG (Khalil and El-Beltagy, 2016) was to use CNN for text classification. For each category, they train a CNN using pre-trained word vectors from Google. The output is given by using result from all 12 classifiers.

- One of the teams BUTknot (Machacek, 2016) used supervised machine learning using bigram bag-of-words model. They did a lot of manual preprocessing including term substitution, lemmatization etc.

- Another team XRCE (Brun et al., 2016) used a 2 step approach. First step classifies the explicit aspect terms detected by CRF into one or more categories. Then in the second step, sentences with NULL aspects are classified into one or more categories.

4.2 Opinion Target Extraction

This task is a sequence llabeling task where we need to mark targets in a sentence.

- NLangp (Toh and Su, 2016) use Conditional Random Fields for this and also uses features from RNN. They use CRFsuite tool (Okazaki, 2007) (Okazaki, 2007) for CRF training. The various features used are Word, Name list, head word, word cluster, dp name list(double propagation), word embeddings. Besides these they also use output probabilties from RNN as additional features.

- AUEB (Xenos et al., 2016) also use sequence labeling using CRF for this task. Features used are Morphological (boolean) feature about the current token- all letters in capital, only digit, capital first letter, existence of punctuation. Lexicon features- POS tags, word affixes, Aspect terms. Other features for unconstrained system includes word embeddings of current words and context.

- UWB (Hercig et al., 2016) also use CRF for this task. They use features such as affixes, bag of bigrams, bag of words filtered by POS, dependency based features, word clusters, cluster bigrams (using the CBOW model computed on the opentable dataset).

4.3 Aspect Sentiment Polarity

For a sentence with multiple opinions, we need to find the context of all the OTE and then assign polarity. Several features used include ngrams, POS tags, dependency based features, in domain sentiment word lists. External resources used were Yelp Dataset, Amazon reviews, TripAdvisor dataset.

- NileTMRG (Khalil and El-Beltagy, 2016) use an ensemble model. This counts votes from three classifiers (positive, negative and neutral). Uses CNN model along with features indicating presence or absence of a certain aspect or domain in a sentence.

- IIT-TUDA (Kumar et al., 2016) make use of lexical expansion for inducing sentiment based words on distributional hypothesis. They created a sentiment lexicon using a seed lexicon and used sum of positive, negative and neutral scores of tokens from induced lexicon. They also use features from other sentiment lexicons such as AFINN (Nielsen, 2011), NRC Hashtag (Kiritchenko et al., 2014), NRC Emotion (Mohammad and Turney, 2013), Bingliu's sentiment lexicon (Hu and Liu, 2004). Also use ngrams and Entity#Attribute pairs as features.

We have described various approaches to different subtasks of ABSA at SemEval. The results of these subtasks are shown in figure 1. In the next section we will describe the literature on Aspect Based Sentiment Analysis apart from the shared task.

5 Aspect Based Sentiment Analysis-Besides SemEval

In this section we describe the general subtasks and approaches to ABSA apart from the SemEval shared task. The task of ABSA includes two major subtasks:

- Aspect Extraction
  This is considered as an information extraction task. We want to find out the target of opinion in our sentence. There can be explicit aspects which are mentioned in the sentence and implicit aspects which are not mentioned in the sentence.
  - For explicit aspects we can have different approaches. One is using the frequent nouns and noun phrases (Hu and Liu, 2004). The assumption is that we are dealing with a single domain. To avoid false positives in this technique, concept of PMI score is used with the meronymy discriminators of that domain and the aspect.
  - Further techniques for finding infrequent aspects make use of dependency relations (Zhuang et al., 2006)
  - Now we get to the supervised learning techniques. We make use of sequence labeling techniques such as Hidden Markov Models (Rabiner and Juang, 1986), Conditional Random Fields (Lafferty et al., 2001). CRFs make use of features such as POS tags, tokens, syntactic dependency, lemmas, ner, etc.
  - Another technique is to make use of topic models where topics are analogous to aspects. However topics can cover both aspects and sentiment words so they need separation. A joint topic model (Mei et al., 2007) was proposed to model both sentiment words and topics.

- Aspect Sentiment Classification
  Now that we are accustomed with the concept of aspects and aspect extraction, we can now delve into the actual task i.e finding sentiment associated with every aspect in the sentence. There are two main approaches that we will discuss now:-
  - Supervised Learning Methods
    The objective is to determine the scope of each sentiment expression.
    * In (Wei and Gulla, 2010) (Wei and Gulla, 2010), a hierarchical classification model was proposed. We still need to find out the scope of the sentiment expression. For this, they have used parsing to determine dependency and other relevant features.
      In (Jiang et al., 2011) (Jiang et al., 2011), a dependency parser was used to generate a set of aspect dependent features for classification. A related approach was to weight feature based on position of the feature relative to target aspect in the pare tree. For comparative sentences, words like than , but and other related words can be used to segment a sentence.
    * As already said, supervised learning requires lot of training data. And also model trained for one domain can perform poorly in other. Domain adaptation is a difficult prob-

Lang./ Dom.	Slot1 F-1	Slot2 F-1	{Slot1,Slot2} F-1	Slot3 Acc.
EN/	NLANG./U/73.031	NLANG./U/72.34	NLANG./U/52.607	XRCE/C/88.126
REST	NileT./U/72.886	AUEB-./U/70.441	XRCE/C/48.891	IIT-T./U/86.729
	BUTkn./U/72.396	UWB/U/67.089	NLANG./C/45.724	NileT./U/85.448
	AUEB-./U/71.537	UWB/C/66.906	TGB/C/43.081*	IHS-R./U/83.935
	BUTkn./C/71.494	GTI/U/66.553	bunji/U/41.113	ECNU/U/83.586
	SYSU/U/70.869	Senti./C/66.545	UWB/C/41.108	AUEB-./U/83.236
	XRCE/C/68.701	bunji/U/64.882	UWB/U/41.088	INSIG./U/82.072
	UWB/U/68.203	NLANG./C/63.861	DMIS/U/39.796	UWB/C/81.839
	INSIG./U/68.108	DMIS/C/63.495	DMIS/C/38.976	UWB/U/81.723
	ESI/U/67.979	XRCE/C/61.98	**basel.**/C/37.795	SeemGo/C/81.141
	UWB/C/67.817	AUEB-./C/61.552	IHS-R./U/35.608	bunji/U/81.024
	GTI/U/67.714	UWate./U/57.067	IHS-R./U/34.864	TGB/C/80.908*
	AUEB-./C/67.35	KnowC./U/56.816*	UWate./U/34.536	ECNU/C/80.559
	NLANG./C/65.563	TGB/C/55.054*	SeemGo/U/30.667	UWate./U/80.326
	LeeHu./C/65.455	BUAP/U/50.253	BUAP/U/18.428	INSIG./C/80.21
	TGB/C/63.919*	**basel.**/C/44.071		DMIS/C/79.977
	IIT-T./U/63.051	IHS-R./U/43.808		DMIS/U/79.627
	DMIS/U/62.583	IIT-T./U/42.603		IHS-R./U/78.696
	DMIS/C/61.754	SeemGo/U/34.332		Senti./U/78.114
	IIT-T./C/61.227			LeeHu./C/78.114
	bunji/U/60.145			**basel.**/C/76.484
	basel./C/59.928			bunji/C/76.251
	UFAL/U/59.3			SeemGo/U/72.992
	INSIG./C/58.303			AKTSKI/U/71.711
	IHS-R./U/55.034			COMMI./C/70.547
	IHS-R./U/53.149			SNLP/U/69.965
	SeemGo/U/50.737			GTI/U/69.965
	UWate./U/49.73			CENNL./C/63.912
	CENNL./C/40.578			BUAP/U/60.885
	BUAP/U/37.29			

Figure 1: Results of different teams for Restaurant Domain and English Language.

lem. So we study this problem in different domains differently.

– Lexicon Based Approaches
We can handle the domain specific issues by incorporating lexicon based approaches. This has been shown to perform quite well in a large number of domains. Such methods are typically unsupervised. The main component of lexicon based approach is the sentiment lexicon. Sentiment lexicon contains Sentiment Words & Phrases, Idioms, Sentiment Shifters, Rules of Opinions, Composite expressions.

Along with the above lexicon features, sentiment shifters play a major role, so they also need to be taken care of. For example, "I like Dominoes Pizza." is a positive review about pizza whereas "I don't like Dominoes Pizza." is completely opposite i.e it is negative. Here "don't" is a sentiment shifter which changes the polarity of the word "like". Now we will discuss one of the methods. This approach is from (Ding, Liu and Yu, 2008) (Ding et al., 2008) and it has four steps and we will use the following example for this method " The voice quality of this phone is not good, but the battery life is long."

1. Mark Sentiment words and phrases- For every sentence that contain one or more aspects, this step marks all sentiment words and phrases. Each positive word is given a +1 score and each negative i given a score of -1. Our sentence now becomes " The voice quality of this phone is not **good**[+1], but the battery life is long."

2. Apply Sentiment Shifters- Sentiment Shifters (also called valence shifters) are the words that can change sentiment orientations. There are several sentiment shifters. Negations words like not, never, none, nobody, nowhere, neither, and cannot are common. Now our sentence becomes " The voice quality of this phone is not good[-1], but the battery life is long."

3. Handle But Clauses- Words and phrases that indicate contrary need to be handled. For example, in the sentence "The voice quality of this phone is not good[-1], but the battery life is long[+1]." due to but close [+1] is added to the end of but clause as the clause before but has [-1]. Contrary to above "but" in "Not only... but also" does not negate the previous sentiment.

4. Aggregate Opinions- This step is used to aggregate all the scores to get the final sentiment score on each aspect. Let the sentence contain the following aspects $(a_1, ...a_n)$and a set of sentiment words $(sw_1....sw_m)$ with their sentiment scores from step 1 to 3. The sentiment score for each aspect is obtained by the following equation -

$$score(a_i, s) = \sum_{ow_j \epsilon s} \frac{sw_j.so}{dist(sw_j, a_i)} \quad (1)$$

This way we get the aggregate sentiment score for every aspect. We can further improve it using dependency based features.

6 Conclusion

ABSA is very important for efficient analysis of opinions and requires multiple tasks to be performed. In this paper, we have explored the task of Aspect Based Sentiment Analysis as a shared task in SemEval- 2016 and also in general. We have described various approaches to these subtasks. Besides this paper, one can search for very recent research work as sentiment analysis is an active research area.

References

Caroline Brun, Julien Perez, and Claude Roux. 2016. Xrce at semeval-2016 task 5: Feedbacked ensemble modeling on syntactico-semantic knowledge for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 277–281. http://www.aclweb.org/anthology/S16-1044.

Alexander Clark. 2000. Inducing syntactic categories by context distribution clustering. In *Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning-Volume 7*. Association for Computational Linguistics, pages 91–94.

Xiaowen Ding, Bing Liu, and Philip S Yu. 2008. A holistic lexicon-based approach to opinion mining. In *Proceedings of the 2008 international conference on web search and data mining*. ACM, pages 231–240.

Tomáš Hercig, Tomáš Brychcín, Lukáš Svoboda, and Michal Konkol. 2016. Uwb at semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 342–349. http://www.aclweb.org/anthology/S16-1055.

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*. ACM, pages 168–177.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, pages 151–160.

Talaat Khalil and Samhaa R. El-Beltagy. 2016. Niletmrg at semeval-2016 task 5: Deep convolutional neural networks for aspect category and sentiment extraction. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 271–276. http://www.aclweb.org/anthology/S16-1043.

Svetlana Kiritchenko, Xiaodan Zhu, and Saif M. Mohammad. 2014. Sentiment analysis of short informal texts. *J. Artif. Intell. Res. (JAIR)* 50:723–762. https://doi.org/10.1613/jair.4272.

Ayush Kumar, Sarah Kohail, Amit Kumar, Asif Ekbal, and Chris Biemann. 2016. Iit-tuda at semeval-2016 task 5: Beyond sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 1129–1135. http://www.aclweb.org/anthology/S16-1174.

John Lafferty, Andrew McCallum, Fernando Pereira, et al. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data .

Jakub Machacek. 2016. Butknot at semeval-2016 task 5: Supervised machine learning with term substitution approach in aspect category detection. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 301–305. http://www.aclweb.org/anthology/S16-1048.

Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*. ACM, pages 171–180.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. Technical report, NRC Technical Report.

F. Å. Nielsen. 2011. Afinn. http://www2.imm.dtu.dk/pubdb/p.php?6010.

Naoaki Okazaki. 2007. Crfsuite: a fast implementation of conditional random fields (crfs). http://www.chokkan.org/software/crfsuite/.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria 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-2016)*. Association for Computational Linguistics, San Diego, California, pages 19–30. http://www.aclweb.org/anthology/S16-1002.

Maria Pontiki, Dimitris 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. http://www.aclweb.org/anthology/S15-2082.

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)*. Association for Computational Linguistics and Dublin City University, Dublin, Ireland, pages 27–35. http://www.aclweb.org/anthology/S14-2004.

Lawrence Rabiner and B Juang. 1986. An introduction to hidden markov models. *ieee assp magazine* 3(1):4–16.

Zhiqiang Toh and Jian Su. 2016. Nlangp at semeval-2016 task 5: Improving aspect based sentiment analysis using neural network features. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 282–288. http://www.aclweb.org/anthology/S16-1045.

Wei Wei and Jon Atle Gulla. 2010. Sentiment learning on product reviews via sentiment ontology tree. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 404–413.

Dionysios Xenos, Panagiotis Theodorakakos, John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos. 2016. Aueb-absa at semeval-2016 task 5: Ensembles of classifiers and embeddings for aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, San Diego, California, pages 312–317. http://www.aclweb.org/anthology/S16-1050.

Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie review mining and summarization. In *Proceedings of the 15th ACM international conference on Information and knowledge management*. ACM, pages 43–50.