Supervised Methods for Aspect-Based Sentiment Analysis

Hussam Hamdan*,,*****
*LSIS
Aix-Marseille Université CNRS
Av. Esc. Normandie Niemen,
13397 Marseille Cedex 20,
France
hussam.hamdan@lsis.org

Patrice Bellot*,**
**OpenEdition
Aix-Marseille Université CNRS
3 pl. V. Hugo, case n°86
13331 Marseille Cedex 3,
France
patrice.bellot@lsis.org

Frederic Béchet***
***LIF
Aix-Marseille Université CNRS
Avenue de Luminy
13288 Marseille Cedex 9,
France
frederic.bechet@lif.univ-mrs.fr

Abstract

In this paper, we present our contribution in SemEval2014 ABSA task, some supervised methods for Aspect-Based Sentiment Analysis of restaurant and laptop reviews are proposed, implemented and evaluated. We focus on determining the aspect terms existing in each sentence, finding out their polarities, detecting the categories of the sentence and the polarity of each category. The evaluation results of our proposed methods exhibit a significant improvement in terms of accuracy and f-measure over all four subtasks regarding to the baseline proposed by SemEval organisers.

1 Introduction

The increasing amount of user-generated textual data has increased the need of efficient techniques for analysing it. Sentiment Analysis (SA) has become more and more interesting since the year 2000 (Liu 2012), many techniques in Natural Language Processing have been used to understand the expressed sentiment on an entity.

Many levels of granularity have been also distinguished: Document Level SA considers the whole document is about an entity and classifies whether the expressed sentiment is positive, negative or neutral; Sentence Level SA determines the sentiment of each sentence, some works have been done on Clause Level SA but they are still not enough; Entity or Aspect-Based SA performs finer-grained analysis in which all entities and

their aspects should be extracted and the sentiment on them should also be determined.

Aspect-Based SA task consists of several subproblems, the document is about many entities which could be for example *a restaurant, a laptop, a printer*. Users may refer to an entity by different writings but normally there are not a lot of variations to indicate the same entity, each entity has many aspects which could be its parts or attributes, some aspects could be another entity such as *screen of laptop*, but most works did not take this case into account. Therefore, we could define the opinion by the quintuple (Liu 2012) (e_i, a_{ij}, s_{ijkl}, h_k, t_l) where e_i is the entity i, a_{ij} are the aspects of the entity i, s_{ijkl} is the expressed sentiment on the aspect at the time t_l, h_k the holder which created the document or the text.

This definition does not take into account that the entity has aspects that could have also other aspects which leads to an aspect hierarchy, in order to avoid this information loss, few works have handled this issue, they proposed to represent the aspect as a tree of aspect terms (Wei and Gulla 2010; Kim, Zhang et al. 2013).

Supervised and unsupervised methods have been used for handling this task, in this paper, we propose supervised methods and test them over two datasets related to laptop reviews and restaurant reviews provided by the ABSA task of SemEval2014 (Pontiki, Galanis et al. 2014). We tackle four subtasks:

1. Aspect term extraction: CRF model is proposed.
2. Aspect Term Polarity Detection: Multinomial Naive-Bayes classifier with some features such as Z-score, POS and prior polarity extracted from Subjectivity

Lexicon (Wilson, Wiebe et al. 2005) and Bing Liu's Opinion Lexicon[1].

3. Category Detection:
   Z-score model for category detection has been used.
4. Category Polarity Detection:
   The same model proposed for aspect term polarity detection has been adopted.

2 Related works

Several methods concerning the ABSA have been proposed, some of them are supervised, and others unsupervised. The earliest work on aspect detection from on-line reviews presented by Hu and Liu used association rule mining based on Apriori algorithm to extract frequent noun phrases as product features, they used two seed sets of 30 positive and negative adjectives, then WordNet has been used to find and add the seed words synonyms. Infrequent aspects had been processed by finding the noun related to an opinionated word (Hu and Liu 2004).

Opinion Digger (Moghaddam and Ester 2010) used also Apriori algorithm to extract the frequent aspects then it filters the non-aspects by applying a constraint -learned from the training data- on the extracted aspects. KNN algorithm is applied to estimate the aspect rating scaling from 1 to 5 stands for (Excellent, Good, Average, Poor, Terrible), assuming that the sentiment is expressed by the nearest adjectives to the aspect term in the sentence segment, WordNet is used for finding the synonyms of sentiment word in order to use them to estimate the distance between it and the words of rating scale.

Some unsupervised methods based on LDA (Latent Dirichlet allocation) were proposed. Brody and Elhadad used LDA to find the aspects, determined the number of topics by applying a clustering method (Brody and Elhadad 2010), then they used a similar method proposed by Hatzivassiloglou and McKeown (Hatzivassiloglou and McKeown 1997) to extract the conjunctive adjectives but not the disjunctive due to the specificity of the domain, seed sets were used and assigned scores, these scores were propagated using propagation method through the aspect-sentiment graph building from the pairs of aspect and related adjectives.

Other works make one LDA based model for the aspect and sentiment extraction. Lin and He

(Lin and He 2009)proposed Joint model of Sentiment and Topic (JST) which extends the state-of-the-art topic model, Latent Dirichlet Allocation (LDA) by adding a sentiment layer, this model is fully unsupervised and it can detect sentiment and topic simultaneously.

Wei and Gulla (Wei and Gulla 2010) modelled the hierarchical relation between product aspects. They defined SOT Sentiment Ontology Tree to formulate the knowledge of hierarchical relationships among product attributes and tackle the problem of sentiment analysis as a hierarchical classification problem. Unsupervised hierarchical aspect

Sentiment model (HASM) was proposed by Kim et al (Kim, Zhang et al. 2013) to discover a hierarchical structure of aspect-based sentiments from unlabelled online reviews.

Supervised methods uses normally a CRF or HMM models. Jin and Ho (Jin and Ho 2009) applied a lexicalized HMM model to extract aspects using the words and their part-of-speech tags in order to learn a model, then unsupervised algorithm for determining the aspect sentiment using the nearest opinion word to the aspect and taking into account the polarity reversal words (such as not). CRF model was used by Jakob and Gurevych (Jakob and Gurevych 2010) with these features: tokens, POS tags, syntactic dependency (if the aspect has a relation with the opinionated word), word distance (the distance between the word in the closest noun phrase and the opinionated word), and opinion sentences (each token in the sentence containing an opinionated expression is labelled by this feature), the input of this method is also the opinionated expressions, they use these expressions for predicting the aspect sentiment using the dependency parsing for retrieving the pair aspect-expression from the training set.

Our method for aspect extraction is closed to (Jakob and Gurevych 2010), where we used CRF model with different features for aspect extraction, but another method for sentiment detection.

The second and fourth subtasks are concerning the polarity detection, so besides to all previous discussed works, we can handle them as sentence level SA. We choose to use Multinomial Naive Bayes with some features (POS, Z-score, pre-polarity). The most related work is (Hamdan, Béchet et al. 2013) where they used Naive Bays with WordNet, DBpedia and SentiWordNet features.

3 The System

Our system is composed of four subtasks:

3.1 Subtask1: Aspect Terms Extraction

The objective of this subtask is to extract all aspect terms in the review sentence, aspect terms could be a word or multiple words. For this purpose we have used CRF (**C**onditional **R**andom **F**ield) which have been used for information extraction. We choose the IOB notation, therefore we distinguish the terms at the **B**eginning, the **I**nside and the **O**utside of aspect term expression. Then, we propose 16 features, for each term we extract the following features:
   - Its root (Porter Stemmer);
   - Its POS tag;
   -The stemming roots for all three words before and after the term;
   -The POS tags for all three words before and after the term;
   - A feature indicates if the word starts with capital letter;
   -A feature indicates if the word is capitalised.
For example, for this review "But the staff was so horrible to us." Where *staff* is the aspect term, the target of each word will be:
 But:O the:O staff:B was:O so:O horrible:O to:O us:O.

3.2 Subtask2: Aspect Term Polarity Detection

This subtask can be seen as sentence level or phrase level sentiment Analysis, the first step (1) we should detect the context or the words related to the aspect term, then to compute its polarity according to these words. Dependency parsing could be used to determine these words or simple distance function. We extract the context of aspect term according to the syntax and other aspect terms. Therefore, the context is the term itself and all the surrounding terms enclosed between two separators (commas in general), if another aspect is also enclosed by these separators we consider it as a separator instead of the comma, and we do not take the terms after it or before it (according to its direction to the aspect term). If the sentence has only an aspect term the separators will be the beginning and the end of the sentence. For example, for this review "*It took half an hour to get our check, which was perfect since we could sit, have drinks and talk!*" where we have two aspect terms *drinks* and *check*, the context of *check* will be "*It took half an hour to get our check*" and the context of *drinks* will be "*have drinks and talk!*". Another example, "*All the money went into the interior decoration, none of it went to the chefs.*" The context for *interior decoration* will be "*All the money went into the interior decoration*" and the context for *chefs* will be "*none of it went to the chefs*".

The second step (2) we should determine the polarity, which could be positive, negative, neutral or conflict. We propose to use Multinomial Naive-Bayes for learning a classifier based on different features:
   - The terms in the sentence (term frequency);
   - The POS features (the number of adjectives, adverbs, verbs, nouns, connectors)
   - The pre-polarity features (the number of positive and negative words in the sentence extracted from Subjectivity lexicon and Bing Liu's Opinion Lexicon);
   - Z-score features (the number of words which have Z-score more than three in each sentiment class), Z_score is described in 3.3.

3.3 Subtask3: Category Detection

Determining the categories of each sentence can be seen as a multi-label classification problem at sentence level.

We propose to use Z-score which is capable of distinguishing the importance of a term in a category. The more the term is important in a category the more its Z-score is high in this category and low in other categories in which it is not important. Thus, we compute the Z-score for all terms using the annotated data, then for each given sentence, the sum of Z-score over each category is computed if the Z-score of term in a category is less than zero, we ignore it in this category because it is not important, the sentence will be attributed to the category having the highest Z-score, if some categories have the same Z-scores the sentence will be attributed to the both. The algorithm steps:
For each tem t in the sentence:
   For each category c:
      If z-score(t,c)>0:
         Z_sc[c]+= z-score(t,c)
Categories=max(Z_sc)

We assume that the term frequency follows the multinomial distribution. Thus, Z_score can be seen as a standardization of the term frequency. We compute Z score for each term t_i in a class C_j (t_{ij}) by calculating its term relative frequency tfr_{ij} in a particular class C_j, as well as the mean

(mean$_i$) which is the term probability over the whole corpus multiplied by n$_j$ the number of terms in the class C$_j$, and standard deviation (sd$_i$) of term ti according to the underlying corpus (see Eq. (1,2)).

$$Z_{score(t_{ij})} = \frac{tfr_{ij} - mean_i}{sdi} \quad Eq. (1)$$

$$Z_{score(t_{ij})} = \frac{tfr_{ij} - n_j * p(ti)}{\sqrt{nj * p(ti) * (1 - p(ti))}} \quad Eq. (2)$$

Z_score was exploited for SA by (Zubaryeva and Savoy 2010), they choose a threshold (Z>2) for selecting the number of terms having Z_score more than the threshold, then they used a logistic regression for combining these scores. We use Z_score as added features for multinomial Naive Bayes classifier.

3.4 Subtask4: Category Polarity Detection

We have used Multinomial Naive-Bayes as in the subtask2 step (2) with the same features, but the different that we add also the name of the category as a feature. Thus, for each sentence having *n* category we add *n* examples to the training set, the difference between them is the feature of the category.

4 Experiments and Evaluations

We tested our system using the training and testing data provided by SemEval 2014 ABSA task. Two data sets were provided; the first contains3Ksentences of restaurant reviews annotated by the aspect terms, their polarities, their categories, the polarities of each category. The second contains of 3K sentences of laptop reviews annotated just by the aspect terms, their polarities.
The evaluation process was done in two steps. First step is concerning the subtasks 1 and 3 which involves the aspect terms extraction and category detection, we were provided with restaurant review and laptop review sentences and we had to extract the aspect terms for both data sets and the categories for the restaurant one. Baseline methods were provided; Table1 demonstrates the results of these subtasks in terms of precision *P*, recall *R* and f-measure *F* for our system and the baseline[2].
We remark that our system is 24% and 21% above the baseline for aspect terms extraction in restaurant and laptop reviews respectively, and

above 3% for category detection in restaurant reviews.

Data	subtask		P	R	F
Res	1	Baseline	0,52	0,42	0,47
		System	**0.81**	**0.63**	**0.71**
	3	Baseline	0,73	0,59	0,65
		System	**0.77**	**0.60**	**0.68**
Lap	1	Baseline	0,44	0,29	0,35
		System	**0.76**	**0.45**	**0.56**

Table 1. Results of subtask 1, 2 for restaurant reviews, subtask 1 for laptop reviews

The second step involves the evaluation of subtask 2 and 4, we were provided with(1) restaurant review sentences annotated by their aspect terms, and categories, we had to determine the polarity for each aspect term and category; (2) laptop review sentences annotated by aspect terms and we had to determine the aspect term polarity. Table 2 demonstrates the results of our system and the baseline (A: accuracy, R: number of true retrieved examples, All: number of all true examples).

Data	subtask		R	All	A
Res	2	Baseline	673	1134	0,64
		System	**818**	**1134**	**0.72**
	4	Baseline	673	1025	0,65
		System	**739**	**1025**	**0.72**
Lap	2	Baseline	336	654	0,51
		System	**424**	654	**0,64**

Table 2. Results of subtask 2, 4 for restaurant reviews, subtask 2 for laptop reviews

We remark that our system is 8% and 13% above the baseline for aspect terms polarity detection in restaurant and laptop reviews respectively, and 7% above for category polarity detection in restaurant reviews.

5 Conclusion

We have built a system for Aspect-Based Sentiment Analysis; we proposed different supervised methods for the four sub-tasks. Our results are always above the baseline proposed by the organiser of SemEval. We proposed to use CRF for aspect term extraction, Z-score model for category detection, Multinomial Naive-Bayes with some new features for polarity detection. We find that the use of Z-score is useful for the category and polarity detection, we are going to test it in another sentiment analysis tasks of another domains.

Reference

Samuel Brody and Noemie Elhadad (2010). An unsupervised aspect-sentiment model for online reviews. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, California, Association for Computational Linguistics**:** 804-812.

Hussam Hamdan,Frederic Béchet and Patrice Bellot (2013). Experiments with DBpedia, WordNet and SentiWordNet as resources for sentiment analysis in micro-blogging. Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA.

Vasileios Hatzivassiloglou and Kathleen R Mckeown (1997). Predicting the semantic orientation of adjectives. Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics.

Minqing Hu and Bing Liu (2004). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. Seattle, WA, USA, ACM**:** 168-177.

Niklas Jakob and Iryna Gurevych (2010). Extracting opinion targets in a single- and cross-domain setting with conditional random fields. Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Cambridge, Massachusetts, Association for Computational Linguistics**:** 1035-1045.

Wei Jin and Hung Hay Ho (2009). A novel lexicalized HMM-based learning framework for web opinion mining. Proceedings of the 26th Annual International Conference on Machine Learning. Montreal, Quebec, Canada, ACM**:** 465-472.

Suin Kim,Jianwen Zhang,Zheng Chen,Alice Oh and Shixia Liu (2013). A Hierarchical Aspect-Sentiment Model for Online Reviews.

Chenghua Lin and Yulan He (2009). Joint sentiment/topic model for sentiment analysis. Proceedings of the 18th ACM conference on Information and knowledge management. Hong Kong, China, ACM**:** 375-384.

Bing Liu (2012). Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers.

Samaneh Moghaddam and Martin Ester (2010). Opinion digger: an unsupervised opinion miner from unstructured product reviews. Proceedings of the 19th ACM international conference on Information and knowledge management. Toronto, ON, Canada, ACM**:** 1825-1828.

Maria Pontiki,Dimitrios Galanis,John Pavlopoulos,Harris Papageorgiou,Ion Androutsopoulos and Suresh Manandhar. (2014). "SemEval-2014 Task 4: Aspect Based Sentiment Analysis." Proceedings of the International Workshop on Semantic Evaluation (SemEval).

Wei Wei and Jon Atle Gulla (2010). Sentiment learning on product reviews via sentiment ontology tree. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden, Association for Computational Linguistics**:** 404-413.

Theresa Wilson,Janyce Wiebe and Paul Hoffmann (2005). Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Vancouver, British Columbia, Canada, Association for Computational Linguistics**:** 347-354.

Olena Zubaryeva and Jacques Savoy (2010). "Opinion Detection by Combining Machine Learning & Linguistic Tools." In Proceedings of the 8th NTCIR, Workshop Meeting on Evaluation of Information Access Technologies: InformationRetrieval, Question Answering and Cross-Lingual Information Access.