

SINAI: Syntactic approach for Aspect Based Sentiment Analysis

Salud M. Jiménez-Zafra, Eugenio Martínez-Cámara,
M. Teresa Martín-Valdivia, L. Alfonso Ureña-López

SINAI Research Group

University of Jaén

E-23071, Jaén (Spain)

{sjzafra, emcamara, maite, laurena}@ujaen.es

Abstract

This paper describes the participation of the SINAI research group in the task Aspect Based Sentiment Analysis of SemEval Workshop 2015 Edition. We propose a syntactic approach for identifying the words that modify each aspect, with the aim of classifying the sentiment expressed towards each attribute of an entity.

1 Introduction

Opinion Mining (OM), also known as Sentiment Analysis (SA), is the discipline that focuses on the computational treatment of opinion, sentiment and subjectivity in texts (Pang and Lee, 2008). Currently, OM is a trendy task in the field of Natural Language Processing, due mainly to the fact of the proliferation of user-generated content and the interest in the knowledge of the opinion of people by consumers and businesses.

Most of the systems developed up to now carry out opinion analysis at document level ((Pang et al., 2002), (Turney, 2002)) or at sentence level ((Wilson et al., 2005), (Yu and Hatzivassiloglou, 2003)), that is, they determine the overall sentiment expressed by the reviewer about the topic, product, person... of study. However, the fact that the overall sentiment of a product is positive does not mean that the author thinks that all the aspects of the product are positives, or the fact that is negative does not involve that everything about the product is bad. For this reason, users and companies are not satisfied with knowing the overall sentiment of a product or service, they

seek a more detailed knowledge. Consequently, to achieve a higher level of detail, part of the scientific community related to this area is working on SA at aspect level ((Quan and Ren, 2014), (Marcheggiani et al., 2014), (Lu et al., 2011), (Thet et al., 2010)) and even, there is a competition on this topic that began to conduct last year (Pontiki et al., 2014) in the International Workshop on Semantic Evaluation 2014 (SemEval 2014).

This year, the 2015 edition of SemEval has also proposed a task for SA at aspect level. The SemEval-2015 Aspect Based Sentiment Analysis task is a continuation of SemEval-2014 Task 4 (Pontiki et al., 2014). The aim of this task is to identify the attributes of an entity that are being reviewed and the sentiment expressed for each one. It is divided into three slots. The first one is focused on the identification of every entity E and attribute A pair (E#A) towards which an opinion is expressed in the given text. Slot 2 proposes to determine the expression used in the text to refer to the reviewed entity, that is, the Opinion Target Expression (OTE). Finally, Slot 3 has as goal to classify the sentiment expressed over each category (E#A pair) as positive, negative or neutral. We have participated in the slot related to sentiment polarity (Slot 3).

Due to the fact that OM is a domain-dependent task, the organization proposes the three slots in different domains, two known (restaurants and laptops) and one unknown until the evaluation (hotels). A wider description of the task and the dataset used can be found in the task description paper (Pontiki et al., 2015).

The rest of the paper is organized as follows. Sec-

tion 2 describes the system developed and the resources that we have used. To sum up, the results reached and an analysis of the same are shown in Section 3.

2 System description - Slot 3

As we have mentioned above, we have taken part in the Slot 3. The aim of this slot is to identify the polarity of each category or each <category, OTE> pair on which an opinion is expressed in a given review. This task has been carried out on two known domains and one unknown domain. For each of the known domains, restaurants and laptops, the organization has provided a dataset for training, whereas for the unknown domain any information has been given until the test set has been released. Therefore, we have used a supervised method for restaurants and laptops domains and we have developed an unsupervised method for the unknown domain.

2.1 Slot 3 - Restaurant domain ABSA

The training data related to restaurants domain contains 254 reviews. Each review is composed of different sentences annotated with opinion tuples. Each opinion tuple has information about the Opinion Target Expression (OTE), the Entity and Attribute pair (E#A category) towards the opinion is expressed, the polarity (positive, negative or neutral) and the position of the OTE in the text (from - to).

Using this information we have developed different experiments for polarity prediction. In all of them an SVM classifier of type C-SVC with linear kernel and the default configuration has been trained, and a 10-fold-cross validation model has been used for the assessment (Table 1).

The features that have provided the best results in the training and that we have used for our participation in this slot are the following. For each <category, OTE, polarity> tuple of the training data, we have used as label the polarity value and as features the words that modify the OTE, their PoS tag, their syntactic relation and their polarity using three lexicons (taking into account negation): SentiWordNet (Baccianella et al., 2010), MPQA (Wilson et al., 2005) and eBLR (enriched version of Bing Liu Lexicon (Hu and Liu, 2004) adapted to restaurant domain). Below, we describe briefly how this infor-

Exp.	Type	Accuracy	Features
Exp_1	U	75.57%	Modifying words, PoS, syntactic relation, polarity (SentiWordNet, MPQA, BinLiu)
Exp_2	U	75.88%	Modifying words, PoS, syntactic relation, polarity (SentiWordNet, MPQA, BinLiu) taking into account negation
Exp_3	U	75.67%	Modifying words, PoS, syntactic relation, polarity (SentiWordNet, MPQA, eBLR)
Exp_4	U	75.94%	Modifying words, PoS, syntactic relation, polarity (SentiWordNet, MPQA, eBLR) taking into account negation

Table 1: Experiments restaurants training data (U = Unconstrained, C = Constrained).

mation has been obtained. Thereby, each <category, OTE> tuple of the test data is classified using its features vector and the trained SVM model.

2.1.1 Features

Words that modify the OTE

We call words that modify an OTE those words that specifically have been used in the review to discuss about the OTE. In order to determine what these words are, we use the Stanford Dependencies Parser¹. This parser was designed to provide a simple description of the grammatical relationships that can appear in a sentence and it can be easily understood and effectively used by people without linguistic expertise who want to extract textual relations (De Marneffe and Manning, 2008). It represents all sentence relationships uniformly as typed depen-

¹<http://nlp.stanford.edu/software/lex-parser.shtml>

dependency relations. In this experiment, we have considered the main relationships for expressing opinion about a noun or nominal expression: using an adjectival modifier (“amod”), an active or passive verb (“nsub”, “nsubjpass”), a noun compound modifier (“nn”) or a dependency relation with another word (“dep”). In this way, for each OTE of a review, we use these relationships to extract all the words that modify the aspect of the entity that has been reviewed and we use them as features. If there is no word related to the aspect using these relationships, the previous word to the OTE and the following four words will be used.

Pos Tag

In addition, for each of the words that modify an aspect we get their particular Part of Speech Tag (noun, verb, adjective...).

Syntactic relations

As it has been mentioned above, the syntactic relation of each modifying word with the OTE has also been used as feature.

Polarity

The last feature of our SVM classifier is the polarity of each modifying word according to three lexicons: SentiWordNet, MPQA and eBLR. In addition, it has been used the fixed window method for the treatment of negation. Then, if any of the preceding or following 3 words is a negative particle (“not”, “n’t”, “no”, “never”...), the modifying word polarity will be reversed (positive \rightarrow negative, negative \rightarrow positive, neutral \rightarrow neutral).

SentiWordNet is a lexical resource that assigns to each synset of WordNet² (Miller, 1995) three sentiment scores (positivity, negativity and objectivity) that describe how positive, negative and objective the terms contained in the synset are.

MPQA is a subjectivity lexicon formed by over 8000 subjectivity clues. For each word, it has information about its prior polarity, its part of speech tag and its grade of subjectivity (strong or weak).

Finally, eBLR is an enriched version of Bing Liu Lexicon that we explain below. As is well-known in the SA research community, the semantic orientation of a word is domain-dependent. Therefore, we decided to generate a list of opinion words for the

²Wordnet is an English lexical database which groups words according to their meaning.

restaurant domain, taking as baseline the Bing Liu Lexicon and using the training data for restaurant domain supplied by the organization. For this, we have employed a corpus-based approach following the methodology of (Molina-González et al., 2013) that consists of the use of a sentiment labeled corpus in order to select the most frequent positive and negative words. A word is added to the list of opinion positive words if it only appears in positive reviews and its frequency exceeds a certain threshold. The same process is followed for negative words. In the case of words that appear in both positive and negative reviews, a word is considered as opinion positive/negative word if the frequency of occurrence in positive/negative reviews exceeds the frequency of occurrence in negative/positive reviews in a certain threshold.

2.2 Slot 3 - Laptops domain ABSA

The training data for laptops domain contains 277 reviews. Each review has different sentences annotated at aspect level with the Entity and Attribute pair (E#A category) towards each opinion is expressed and the polarity (positive, negative or neutral). In this case no information about the OTE is provided and thus, we have followed a different approach to that used in the restaurant domain. We have also developed different experiments with an SVM classifier of type C-SVC with linear kernel and the default configuration, and we have also used a 10-fold-cross validation model for the assessment but with different features (Table 2).

Exp.	Type	Accuracy	Features
Exp_1	C	75.08%	Unigrams, PoS
Exp_2	U	73.76%	Unigrams, total positive words (Bin Liu), total negative words (Bin Liu)
Exp_3	U	79.64%	Unigrams, total positive words (eBLL), total negative words (eBLL)

Table 2: Experiments laptops training data (U = Unconstrained, C = Constrained).

For this domain we have submitted two runs, one constrained (using only the provided training data) and another unconstrained (using additional resources for training). These experiments are those that have provided better results with the training data and we have used them for our participation in this domain.

- SINAI_B_Lap_1 (Exp.1 - constrained). For each $\langle \text{category}(E\#A \text{ pair}), \text{polarity} \rangle$ tuple of the training data we have used as label the polarity and as features the entity and the specific attribute of this entity about someone is reviewing, and all the words of the sentence with their pertinent Part of Speech Tag.
- SINAI_B_Lap_2 (Exp.3 - unconstrained). In this case, the features that we have selected for each $\langle \text{category}(E\#A \text{ pair}), \text{polarity} \rangle$ tuple of the training data are the entity and the attribute about someone is reviewing, all the words of the sentence and the number of positive and negative opinion words according to eBLL. eBLL is an enriched version of Bing Liu Lexicon for laptops domain. It has been built using the training data supplied by the organization for laptops domain, in the same way that eBLR Lexicon.

Thus, given a category of the test data, it is classified using its features vector and the trained SVM model.

2.3 Slot 3 - Out of domain ABSA

For the last domain, the organization has not provided any information until the test set has been released. We only knew that we had to assign a polarity value for each $\langle \text{OTE}, \text{category} \rangle$ tuple present in the test data. In this case we have followed an unsupervised approach that we present below.

In order to classify the sentiment expressed about each OTE is important to determine the words that have been used in the review to discuss about the aspect. For this, we have employed the Stanford Dependencies Parser and the main relationships for expressing opinion about a noun or nominal expression: “amod”, “nsubj”, “nsubjpass”, “nn”, “dep” (they are explained in Subsection 2.1). In this way, for each OTE of a review, we use these relationships

to extract all the words that modify it and we use them to determine the sentiment expressed about the OTE. If there is no word related to the aspect using these relationships, the previous word to the aspect and the following four words will be used. We calculate the polarity of each OTE through a voting system based on three classifiers: Bing Liu Lexicon, SentiWordNet and MPQA. To do this we determine, with each of the classifiers individually, the polarity of an OTE using the words that modify it. Thus, according to Bing Liu Lexicon, we count the number of positive (pw) and negative words (nw) that modify the OTE and tag it following the equation 1. On the other hand, we use MPQA as classifier following the same approach but in this case we take into account the PoS of the modifying words in order to get their polarity. At last, we employ SentiWordNet also following the approach of comparing the number of positive and negative words but as this lexicon assigns three sentiment scores to each synset, we calculate the polarity of each modifying word using the Denecke method (Denecke, 2008), that is, we calculate the average of the positivity, negativity and objectivity scores of all the synsets of the word with the same PoS and assign the word the polarity of the highest average.

$$pol(OTE) = \begin{cases} positive & \text{if } (pw > nw) \\ negative & \text{if } (pw < nw) \\ neutral & \text{if } (pw = nw) \end{cases} \quad (1)$$

Therefore, an OTE is positive/negative if there are at least two classifiers that tag it as positive/negative and neutral in another case. It may happen that an OTE is affected by negation, so if any of the preceding or following 3 words is a negative particle (“not”, “n’t”, “no”, “never”...), the OTE polarity will be reversed (positive \rightarrow negative, negative \rightarrow positive, neutral \rightarrow neutral).

3 Analysis of results

This section shows the results reached in the evaluation of the task using the system described in Section 2. Table 3 presents the official results of our submissions. We also include the results of the best team and the average of all participants for comparison.

A clear difference between the results obtained by our team and the average may be seen in Table 3.

Furthermore, the results in restaurants and laptops domain are worse than those achieved in the training phase (Table 1 and Table 2). Therefore, we have calculated the confusion matrix related to each experiment for a deeper analysis (Table 4, Table 5, Table 6 and Table 7).

	Accuracy		
	SINAI	Avg.	Best team
Restaurants	0.6071 (U)	0.7119	0.7870 (U)
Laptops	0.6586 (C) 0.5184 (U)	0.7093	0.7935 (U)
Hotels	0.6372 (U)	0.7079	0.8053 (U)

Table 3: Results test data (U = Unconstrained, C = Constrained).

	Restaurants			
	Pred. pos.	Pred. neu.	Pred. neg.	Recall
Real pos.	446	0	8	0.9824
Real neu.	43	0	2	0
Real neg.	276	3	67	0.1936
Precision	0.583	0	0.8701	

Table 4: Confusion matrix restaurants submission.

	Laptops (C)			
	Pred. pos.	Pred. neu.	Pred. neg.	Recall
Real pos.	491	0	50	0.9076
Real neu.	51	0	28	0
Real neg.	195	0	134	0.4073
Precision	0.6662	0	0.6321	

Table 5: Confusion matrix laptops constraint submission.

	Laptops (U)			
	Pred. pos.	Pred. neu.	Pred. neg.	Recall
Real pos.	391	0	150	0.7227
Real neu.	63	0	16	0
Real neg.	228	0	101	0.3070
Precision	0.5733	0	0.3783	

Table 6: Confusion matrix laptops unconstraint submission.

	Hotels			
	Pred. pos.	Pred. neu.	Pred. neg.	Recall
Real pos.	181	56	6	0.7449
Real neu.	5	6	1	0.5
Real neg.	15	40	29	0.3452
Precision	0.9005	0.0588	0.8056	

Table 7: Confusion matrix hotels submission.

	Restaurants	Laptops
Positive opinions	1198	1103
Neutral opinions	53	106
Negative opinions	403	765

Table 8: Opinions in training data per class.

If we observe Table 4, Table 5 and Table 6, we can see that, in restaurants and laptops domains, the system has failed mainly in the classification of negative and neutral opinions. It has classified most of them as positive. We think that one of the reasons may be that the training data for restaurants and laptops domains is unbalanced (Table 8). For restaurants, the number of positive opinions is almost three times the number of negative opinions. Another possible reason, in restaurants domain, is that we have only taken into account the scope (words that modify the OTE) and not the whole context (all words present in the review). In future works, we will do experiments balancing the datasets in order to test how the system works. Furthermore, we will take into account the whole context in restaurants domain to see if that improves the system.

Regarding the unsupervised system, that has been

tested with hotels domain, there are also differences with respect to the mean accuracy of all teams (Table 3). This is a first approach that can be improved with the consideration of other relationships to determine which words modify the OTE and with a treatment of negation more exhaustive. In future works we will consider these possible improvements.

Acknowledgments

This work has been partially supported by a grant from the Fondo Europeo de Desarrollo Regional (FEDER), ATTOS project (TIN2012-38536-C03-0) from the Spanish Government, AORESCU project (P11-TIC-7684 MO) from the regional government of Junta de Andalucía and CEATIC-2013-01 project from the University of Jaén.

References

- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.
- Marie-Catherine De Marneffe and Christopher D Manning. 2008. Stanford typed dependencies manual.
- Kerstin Denecke. 2008. Using sentiwordnet for multilingual sentiment analysis. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, pages 507–512. IEEE.
- 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*, pages 168–177. ACM.
- Bin Lu, Myle Ott, Claire Cardie, and Benjamin K Tsou. 2011. Multi-aspect sentiment analysis with topic models. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*, pages 81–88. IEEE.
- Diego Marcheggiani, Oscar Täckström, Andrea Esuli, and Fabrizio Sebastiani. 2014. Hierarchical multi-label conditional random fields for aspect-oriented opinion mining. In *Advances in Information Retrieval*, pages 273–285. Springer.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- M Dolores Molina-González, Eugenio Martínez-Cámara, María-Teresa Martín-Valdivia, and José M Perea-Ortega. 2013. Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, 40(18):7250–7257.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10, EMNLP '02*, pages 79–86, Stroudsburg, PA, USA.
- Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, 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.
- Maria Pontiki, Dimitrios Galanis, Harris 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), Denver, Colorado*.
- Changqin Quan and Fuji Ren. 2014. Unsupervised product feature extraction for feature-oriented opinion determination. *Information Sciences*, 272:16–28.
- Tun Thura Thet, Jin-Cheon Na, and Christopher SG Khoo. 2010. Aspect-based sentiment analysis of movie reviews on discussion boards. *Journal of Information Science*, page 0165551510388123.
- Peter D. Turney. 2002. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02*, pages 417–424, Stroudsburg, PA, USA.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, pages 347–354, Stroudsburg, PA, USA.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP '03*, pages 129–136, Stroudsburg, PA, USA.