# Lsislif: CRF and Logistic Regression for Opinion Target Extraction and Sentiment Polarity Analysis

**Hussam Hamdan**
Aix-Marseille University
hussam.hamdan@lsis.org

**Patrice Bellot**
Aix-Marseille University
patrice.bellot@lsis.org

**Frederic Bechet**
Aix-Marseille University
frederic.bechet@lif.univ-mrs.fr

## Abstract

This paper describes our contribution in Opinion Target Extraction OTE and Sentiment Polarity sub tasks of SemEval 2015 ABSA task. A CRF model with IOB notation has been adopted for OTE with several groups of features including syntactic, lexical, semantic, sentiment lexicon features. Our submission for OTE is ranked fifth over twenty submissions. A Logistic Regression model with a weighting schema of positive and negative labels have been used for sentiment polarity; several groups of features (lexical, syntactic, semantic, lexicon and Z score) are extracted. Our submission for Sentiment Polarity is ranked third over ten submissions on the restaurant data set, third over thirteen on the laptops data set, but the first over eleven on the hotel data set that is out-of-domain set.

## 1 Introduction

Sentiment Analysis (SA) has become more and more interesting since the year 2000, 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 papers have focused 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 towards 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 work 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 work has handled this issue, they proposed to represent the aspect as a tree of aspect terms.

In this paper, we focus on Opinion Target Extraction (OTE) and Sentiment Polarity towards a target or a category. The description of each subtask is provided by ABSA organizers (Pontiki et al., 2015). For OTE or aspect term extraction, a CRF model is proposed with IOB annotation and several groups of features including syntactic, lexical, semantic, sentiment lexicon features. For aspect term polarity detection, a logistic regression classifier is trained with weighting schema for positive and negative labels and several groups of features are extracted includ-

753

ing lexical, syntactic, semantic, lexicon and Z score features.

The rest of this paper is organized as follows. Section 2 outlines existing work in aspect extraction and polarity detection. Section 3 describes our system for aspect term extraction. Aspect term polarity detection is presented in Section 4. Section 6 shows the conclusion and the future work.

## 2 Related Work

Aspect-Based Sentiment Analysis consists of several sub tasks. Some papers have proposed different methods for aspect detection and sentiment polarity analysis, others have proposed joint models in order to obtain the aspect and their sentiments from the same model, these models are generally unsupervised or semi-supervised.

The earliest work on aspect detection from online reviews presented by Hu and Liu (Hu and Liu, 2004) that used association rule mining based on Apriori algorithm to extract frequent noun phrases as product features, for polarity detection they used two seed sets of 30 positive and negative adjectives, then WordNet has been used to find and add the synonyms of the seed words. Infrequent features had been processed by finding the noun related to an opinionated word.

Opinion Digger (Moghaddam and Ester, 2010) also used Apriori algorithm to extract the frequent aspects. KNN algorithm is applied to estimate the aspect rating scaling from 1 to 5 stands for (Excellent, Good, Average, Poor, Terrible).

Supervised methods uses normally the 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). A CRF model was used by Jakob and Gurevych (Jakob and Gurevych, 2010) with the following 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 labeled 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. A CRF model is also used by (Hamdan et al., 2014b) with lexical and POS features.

Unsupervised methods based on LDA (Latent Dirichlet allocation) have been proposed. Brody and Elhadad (Brody and Elhadad, 2010) used LDA to figure ou the aspects, determined the number of topics by applying a clustering method, 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. Lin and He (Lin et al., 2012) proposed Joint model of Sentiment and Topic (JST) which extends the state-of-the-art topic model (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) modeled the hierarchical relation between product aspects. They defined Sentiment Ontology Tree (SOT) to formulate the knowledge of hierarchical relationships among product attributes and tackled the problem of sentiment analysis as a hierarchical classification problem. Unsupervised hierarchical aspect Sentiment model (HASM) was proposed by Kim et al (Kim et al., 3 07) to discover a hierarchical structure of aspect-based sentiments from unlabeled online reviews.

Aspect term polarity detection can be seen as a sentence level sentiment analysis. Therefore, many papers can be mentioned. Supervised methods have been widely exploited for this purpose, a classification algorithms with a wise feature extraction could achieve good results (Mohammad et al., 2013) (Hamdan et al., 2015a) (Hamdan et al., 4 29).

## 3 Opinion Target Expression (OTE)

An opinion target expression (OTE) is an expression used in the given text to refer to an aspect or

an aspect term related to the reviewed entity. The objective of OTE slot is to extract all opinion target expressions in a restaurant review, OTE could be a word or multiple words. For this purpose, we have used CRF (Conditional Random Field) which have proved its performance in information extraction. We choose the IOB notation for representing each sentence in the review. Therefore, we distinguish the terms at the Beginning, the Inside and the Outside of OTE. For example, for this review *"But the staff was so horrible to us."* Where *staff* is OTE, the target of each word will be:

But:O the:O staff:B was:O so:O horrible:O to:O us:O.

We extract for each single word the following features for the word itself and the 2 and 3 previous and subsequent words, respectively.

-word lemma using WordNet.

-word POS using NLTK parser.

-word shape: the shape of each character in the word (capital letter, small letter, digit, punctuation, other symbol)

-word type: the type of the word (uppercase, digit, symbol, combination )

-Named entity: the IOB annotation for the named entity extracted from the review using Senna (Collobert, 2011).

-chunk: the chunk of the word (NP, VP, PP) extracted using Senna.

-polarity: the sum of word polarity score calculated using Bing Liu Lexicon (Hu and Liu, 2004) and MPQA subjectivity Lexicon (Wilson et al., 2005).

-Prefixes (all prefixes having length between one to four ).

-Suffixes (all suffixes having length between one to four).

-Stop word: if the word is a stop word or not.

-if the initial letter is uppercase, if all letters are uppercase, All letters lowercase, All letters digit, Contains a uppercase letter, Contains a lowercase letter, Contains a digit, Contains a alphabet letter, Contains a symbol.

We also extract the value of each two successive features in the the range -2,2 (the previous and subsequent two words of actual word) for the following features:

word surface, word POS, word chunk, word shape, word type.

Finally, we extract the value of each three successive features in the the range -1,1 for the two features: word POS and word lemma.

## 3.1 Experiments

The data set is extracted from restaurant reviews, provided by SemEval 2015 ABSA organizers (Pontiki et al., 2015). Table 1 shows the training and testing data sets statistics of restaurant reviews, where each review is composed of several sentences and each sentence may contain several OTE. CRFsuite tool is used for this experiment with lbfgs algorithm. This tool is fast in training and tagging (Okazaki, 2007).

| Data | Reviews | Sentences | OTE |
|------|---------|-----------|------|
| Train | 254 | 1315 | 1654 |
| Test | 96 | 685 | 845 |

Table 1. Training and testing data sets for restaurant OTE slot.

Our submission is ranked fifth with the F1 score over twenty submissions with gain of 14% over the baseline provided by the organizers. This baseline uses the training reviews to create for each category *c* a list of targets to which it is linked to. Then, given a test sentence *s* and a category *c*, the baseline finds the first occurrence in s of each target encountered in *c*s list. Table 2 shows our system and the baseline results.

| Experiment | Recall | Precision | F1 Score |
|------------|--------|-----------|----------|
| Our System | 0.55 | 0.72 | 0.62 |
| Baseline | - | - | 0.48 |

Table 2. The results of OTE slot.

## 4  Sentiment Polarity

For a given set of aspect terms within a sentence, we determine whether the polarity of each aspect term is positive, negative, neutral. For example, the system should extract the polarity of `fajitas` and *salads* in the following sentence: *"I hated their fajitas, but their salads were great"*, fajitas: negative and salads: positive.

This sub-task can be seen as sentence level or phrase level sentiment Analysis. At the first step, we detect the context of the aspect term or OTE,

755

the context is the aspect term itself and all the surrounding terms enclosed between two separators like (,, ;, !), if another aspect term is also enclosed by these separators we consider it as a separator instead, and we do not take the terms after it or before it (according to its direction to the current 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"*.

At the second step, we should determine the polarity, which could be positive, negative, neutral. We propose to use a logistic regression classifier with weighting schema of positive and negative labels with the following features:

**- Word n-grams Features**

Unigrams and bigrams are extracted for each word in the context without any stemming or stop-word removing, all terms with occurrence less than 3 are removed from the feature space.

**- Sentiment Lexicon-based Features**

The system extracts four features from the manual constructed lexicons (Bing Liu Lexicon (Hu and Liu, 2004) and MPQA subjectivity Lexicon (Wilson et al., 2005)) and six features from the automatic ones (NRC Hashtag Sentiment Lexicon (Mohammad, 6 07), Sentiment140 Lexicon (Mohammad et al., 2013), and SentiWordNet (Baccianella et al., 2010)). For each context the number of positive words, the number of negative ones, the number of positive words divided by the number of negative ones and the polarity of the last word are extracted from manual constructed lexicons. In addition to the sum of the positive scores and the sum of the negative scores from the automatic constructed

lexicons.

**- Negation Features**

The rule-based algorithm presented in Christopher Potts Sentiment Symposium Tutorial is implemented. This algorithm appends a negation suffix to all words that appear within a negation scope which is determined by the negation key and a certain punctuation. All these words are added to the feature space.

**4- Z score Features**

Z score can distinguish the importance of each term in each class, their performances have been proved (Hamdan et al., 2014a). We assume as in the mentioned work that the term frequencies are following the multi-nomial distribution. Thus, Z score can be seen as a standardization of the term frequency using multi-nomial distribution. We compute the Z score for each term ti 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). We tested different threshold for choosing the words which have higher Z score, we found 3 is the best one for restaurant data and 4 for laptop data.

$$Zscore(ti) = \frac{tfr_{ij} - mean_i}{sd_i} \qquad (1)$$

Thus, we added the number of words having Z score higher than 3,4 in each class positive,negative and neutral, the two classes which have the maximum number and minimum numbers of words having Z score higher than the threshold. These 5 features have been added to the feature space.

**- Brown Cluster Features**

Each word in the text is mapped to its cluster in Brown clusters, 1000 features are added to feature space where each feature represents the number of words in the text mapped to each cluster. The 1000 clusters is provided in Twitter Word Clusters of CMU ARK group which were constructed from approximately 56 million tweets.

**- Category Feature**

We also added the category of each OTE as a feature to the feature space.

## 4.1 Experiments

In addition to the restaurant data set presented in tabel 1, two other data sets statistics are presented in table 3 (Laptops data which consists of training and testing data sets while the Hotel test set is out of domain set that was provided to test our model on new domain without having training data).

We trained a L1-regularized Logistic regression classifier implemented in LIBLINEAR, which has given good results in several papers (Hamdan et al., 2015b) (Hamdan et al., 2015a). The classifier is trained on the training data set using the previous features with the three polarities (positive, negative, and neutral) as labels. A weighting schema is adapted for each class, we use the weighting option -$wi$ which enables a use of different cost parameter $C$ for different classes. Since the training data is unbalanced, this weighting schema adjusts the probability of each label. Thus, we tuned the classifier in adjusting the cost parameter $C$ of Logistic Regression, weight $w_{pos}$ of positive class and weight $w_{neg}$ of negative class.

We used the 1/10 of training data set for tuning the three parameters in the two data sets (Restaurant, Laptop), all combinations of C in range 0.1 to to 4 by step 0.1, $w_{pos}$ in range 1 to 8 by step 0.1, $w_{neg}$ in range 1 to 8 by step 0.1 are tested. The combination $C$=0.3, textit$w_{pos}$=1.2, $w_{neg}$=1.9 have been chosen for the restaurant set and $C$=0.2 $w_{pos}$=2.1 $w_{neg}$=1.9 for the laptops set.

| Data | Reviews | Sentences | OTE |
|---|---|---|---|
| Train Lap | 277 | 1739 | 1973 |
| Test Lap | 173 | 761 | 949 |
| Test hotel | 30 | 266 | 339 |

Table 3. Data set statistics for Hotel and Laptops Reviews.

Table 4 shows the results of our system on the three data sets. It should note that we use the trained classifier on restaurant data set for predicting the polarity in the Hotel test set the out-of-domain set. Our system outperforms the baseline over the three data set. The gain is of 11.95%, 7.9%, 14.16% in

restaurant, laptop, hotel reviews respectively. The baseline of Hotels is the majority baseline while the other baselines are provide by the organizers which use a trained SVM on the BOW features and the category name feature in each data set. Our system is ranked third over ten submissions in the restaurant data set, third over thirteen in the laptops set, and the first over eleven in the hotel set.

| Experiment | Correct | All | Accuracy |
|---|---|---|---|
| **Restaurant** | | | |
| Our system | 638 | 845 | 75.5 |
| Baseline | 537 | 845 | 63.55 |
| **Laptops** | | | |
| Our system | 739 | 949 | 77.87 |
| Baseline | 664 | 949 | 69.97 |
| **Hotels** | | | |
| Our system | 291 | 339 | 85.84 |
| Baseline | 243 | 339 | 71.68 |

Table 4. Results of sentiment polarity in Restaurant, laptops, hotels reviews.

## 5 Conclusion and future work

We have built two systems for opinion target extraction of restaurant data set, and sentiment polarity analysis for three data sets (restaurant and laptops) and one out-of-domain set (hotel). We have used supervised tagger for OTE, trained a CRF model with several features. A Logistic regression classifier is used for sentiment polarity where we adopted a weighting schema in each domain and applied the same classifier and weighting schema trained on restaurant set on the Hotel test set. In future work, we will focus on using parsing tree for determining the context of OTE instead of the syntactic method. And play with other types of features for the two subtasks OTE and Sentiment Polarity.

## References

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *in Proc. of LREC*.

Brody, S. and Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. In *Human Language Technologies: The 2010 Annual Conference*

*of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 804–812. Association for Computational Linguistics.

Collobert, R. (2011). Deep learning for efficient discriminative parsing. In Gordon, G. J. and Dunson, D. B., editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, volume 15, pages 224–232. Journal of Machine Learning Research - Workshop and Conference Proceedings.

Hamdan, H., Bechet, F., and Bellot, P. (2013-04-29). Experiments with DBpedia, WordNet and SentiWordNet as resources for sentiment analysis in micro-blogging. In *International Workshop on Semantic Evaluation SemEval-2013 (NAACL Workshop)*.

Hamdan, H., Bellot, P., and Bechet, F. (2014a). The impact of z_score on twitter sentiment analysis. In *In Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval 2014)*, page 636.

Hamdan, H., Bellot, P., and Bechet, F. (2014b). Supervised methods for aspect-based sentiment analysis. In *In Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval 2014)*.

Hamdan, H., Bellot, P., and Bechet, F. (2015a). lsislif: Feature extraction and label weighting for sentiment analysis in twitter. In *In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.

Hamdan, H., Bellot, P., and Bechet, F. (2015b). Sentiment lexicon-based features for sentiment analysis in short text. In *In Proceeding of the 16th International Conference on Intelligent Text Processing and Computational Linguistics*.

Hatzivassiloglou, V. and McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the Eighth Conference on European Chapter of the Association for Computational Linguistics*, EACL '97, pages 174–181. Association for Computational Linguistics.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177. ACM.

Jakob, N. and Gurevych, I. (2010). Extracting opinion targets in a single- and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 1035–1045. Association for Computational Linguistics.

Jin, W. and Ho, H. H. (2009). A novel lexicalized HMM-based learning framework for web opinion mining-NOTE FROM ACM: A joint ACM conference committee has determined that the authors of this article violated ACM's publication policy on simultaneous submissions. therefore ACM has shut off access to this paper. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 465–472. ACM.

Kim, S., Zhang, J., Chen, Z., Oh, A., and Liu, S. (2013-07). A hierarchical aspect-sentiment model for online reviews. In *Proceedings of The Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13)*. AAAI.

Lin, C., He, Y., Everson, R., and Ruger, S. (2012). Weakly supervised joint sentiment-topic detection from text. 24(6):1134–1145.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.

Moghaddam, S. and Ester, M. (2010). Opinion digger: An unsupervised opinion miner from unstructured product reviews. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1825–1828. ACM.

Mohammad, S. (2012-06-07). #emotional tweets. In *\*SEM 2012: The First Joint Conference on Lexical and Computational Semantics Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 246–255. Association for Computational Linguistics.

Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). NRCCanada: Building the state-of-the-art in sentiment analysis of tweets. In *In Proceedings of the International Workshop on Semantic Evaluation, SemEval 13*.

Okazaki, N. (2007). *CRFsuite: a fast implementation of Conditional Random Fields (CRFs)*.

Pontiki, M., Galanis, D., Papageogiou, H., Manandhar, S., and Androutsopoulos, I. (2015). SemEval-2015 task 12: Aspect based sentiment analysis. In *In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.

Wei, W. and Gulla, J. A. (2010). Sentiment learning on product reviews via sentiment ontology tree. In *In Proceedings of the 48th Annual Meeting of the ACL*, pages 404–413. ACL.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. (2005). OpinionFinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, HLT-Demo '05, pages 34–35. Association for Computational Linguistics.