

# XRCE at SemEval-2016 Task 5: Feedbacked Ensemble Modelling on Syntactico-Semantic Knowledge for Aspect Based Sentiment Analysis

Caroline Brun and Julien Perez and Claude Roux

Xerox Research Centre Europe

6, chemin de Maupertuis

38240 Meylan, France

{caroline.brun, julien.perez, claude.roux}@xrce.xerox.com

## Abstract

This paper presents our contribution to the SemEval 2016 task 5: Aspect-Based Sentiment Analysis. We have addressed Subtask 1 for the restaurant domain, in English and French, which implies opinion target expression detection, aspect category and polarity classification. We describe the different components of the system, based on composite models combining sophisticated linguistic features with Machine Learning algorithms, and report the results obtained for both languages.

## 1 Introduction and Related Work

Sentiment Analysis is an important topic in natural language processing, and Aspect Based Sentiment Analysis (ABSA), i.e. detection of sentiments expressed on different aspects of a given entity, constitute a very interesting but quite challenging task (Liu, 2012; Ganu et al., 2009). ABSA is a task first introduced at SemEval in 2014 (Pontiki et al., 2014), continued in 2015 (Pontiki et al., 2015) and now, in 2016 (Pontiki et al., 2016). Our team has participated to the first edition, with good results on the restaurant domain (Brun et al., 2014) and decided to reiterate the participation in 2016, on the same domain but on English and French, as the challenge has become multilingual. While relatively similar, the task has evolved since 2014: aspect targets and categories are annotated together instead of separately; only opinionated terms (Opinion Target Expressions, OTE) are annotated, and aspect categories are finer grained (12 classes instead of 5), which makes the subtasks even more challenging.

In the previous challenges, most systems, including ours, use state-of-the art machine learning algorithms such as SVMs (Wagner et al., 2014; Kiritchenko et al., 2014; Brun et al., 2014; Brychcín et al., 2014) or CRFs (Toh and Wang, 2014; Hamdan et al., 2015), with lexical information, bigrams and POS as features. In 2014, (Kiritchenko et al., 2014) had particularly good results on aspect category and aspect polarity detection, using SVMs combined with rich linguistic features including dependency parsing. In 2015, the system presented by (Saías, 2015) reported the best result for polarity classification, using a maximum entropy classifier, having bag-of-words, lemmas, bigrams after verbs, and punctuation based features, along with sentiment lexicon-based features. Our system shares with these ones the use of syntactic features to address the different ABSA tasks.

For the present challenge, we addressed subtask 1, which implies target terms detection, aspect category and polarity classification. In the remaining of the paper, we describe the different components of our system which combine rich linguistic features and machine learning algorithms (CRF, Ensemble models for classification). We then present and discuss our results on the different subtasks. We finally conclude and propose future directions.

## 2 System Description

We present here the different components of our system, dedicated to linguistic feature extraction, sequence labeling and classification.

## 2.1 Linguistic Feature Factory

We use a robust syntactic parser (XIP (Ait-Mokhtar et al., 2002)) as one of the fundamental components of our system. This parser provides a full processing chain including tokenization, morpho-syntactic analysis, POS tagging, Named Entity Detection, chunking and finally, extraction of dependency relations such as subject, object and modifiers. This robust parser has been already used for the Aspect Based Sentiment Analysis of SemEval 14 (Brun et al., 2014). We have designed and adapted a semantic extraction component that extracts semantic information about aspect targets and their polarities on top of the parser described before. For this task, syntactic dependencies, lexical information about word polarities and semantic classes, sub-categorization information are all combined within the parser to extract semantic relations associated to aspect targets. We already have developed a component that extracts sentiment relations (see (Brun, 2011) for the complete description of this component), taking into account contexts and scope of the polar predicates. This semantic component makes use of a polar lexicon associating polarities (only positive and negative) to words, and a semantic lexicon associating lexical semantic features (FOOD, DRINK, AMBIENCE, SERVICE, RESTAURANT, PRICE, STYLE) to words. For the present challenge, we used the English and French version of the grammars, and complemented the existing domain lexicons with lexical information extracted from the training corpus.

We use this parser as a “feature factory”, that outputs linguistic features which can feed the various Machine Learning algorithms we applied to the ABSA tasks<sup>1</sup>, that are described afterwards.

## 2.2 Domain Term detection using CRF

Conditional Random Fields (CRF) (Lafferty et al., 2001) is a popular class of statistical modelling for sequence labeling, which can be applied to term detection. For this specific task, we used CRFsuite<sup>2</sup> (Okazaki, 2007), which provides a cross-validation

<sup>1</sup>**Important note:** as we use this proprietary parser for our experiments, and for reproducibility concerns, we have released our feature datasets both for train and test, on English and French, for the different slots at this url: <http://semeval2016.xrce.xerox.com/>

<sup>2</sup><http://www.chokkan.org/software/crfsuite>

mechanism, to detect the terms and categories in the different sentences. The CRF model was trained with some traditional features such as POS, lemma, surface form, and the presence of uppercase letters for instance. However, we also used the output of the XIP parser to detect if a word was in a particular syntactic construction such as attribute, coordination or object dependencies. The parser also supplies lexico-semantic information, for instance if a given noun phrase is related to food or to service, also integrated into the list of features. Finally, we used as feature whether a word was detected as being part of a sentiment analysis structure. We then combined the features from the three previous and the three next words in order to train our system within a window of seven words. Thus, the CRF model was trained over a mixture of word forms and syntax. Since CRFsuite provides a cross-validation scheme, we applied a 10 fold cross validation, which displayed a consistent F-measure of 85 over the English training set and the French training set.

## 2.3 Inference models

In this section, the different aspects of the decision model are presented. Based on the rich linguistic representation produced for the different tasks, a feedbacked loop of classification has been developed. Indeed, in order to highlight the most efficient linguistic features, we developed a simple framework of feedback generation for assisting feature definition and selection.

### 2.3.1 Feedbacked Ensemble Models

In order to address aspect category and polarity classification, an interactive feedbacked ensemble method pipeline has been designed to cope with the strong sparsity nature of the data. Indeed, figure 1 details the overall dynamic of the model. First, the feature set associated with the considered term is defined. Then, in order to cope with sparsity truncated singular value decomposition (Hansen, 1986) can be performed on the original set of features then a one-versus-all Elastic Net regression model (Zou and Hastie, 2003) is used to infer the target concept, in our case category and polarity. The advantage of Elastic Net is that it explicitly defines a trade-off between L1-norm and L2-norm type of regularization. As an output of the model learning task, a model rep-

representation and cross-validation scores are provided in order to allow for improvement of the feature set used as decision support, enabling a formal error analysis of such model. As feedbacked interaction, statistics informing of the relevance of the sentence features estimated during crossvalidation but also recurrent errors occurring in crossvalidation have been used as evidence in order to enhanced the linguistic representation of the sentences.

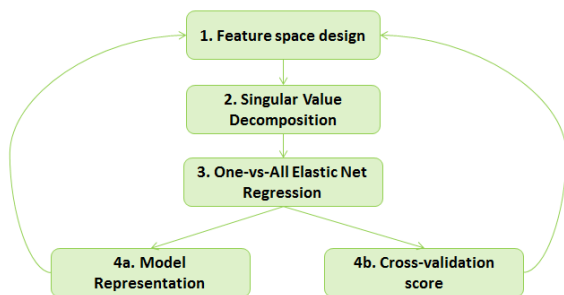


Figure 1: Ensemble Modeling Process for ABSA

This pipeline is applied to the different classification tasks described below.

### 2.3.2 Aspect Category Classification

For the restaurant domain, 12 semantic categories are covering the aspects (food#quality, food#style\_options, food#prices, drinks#quality, drinks#style\_options, drinks#prices, location#general, restaurant#general, restaurant#misc, restaurant#prices, service#general and ambience#general), into which explicit and implicit aspect targets have to be classified. Classification into aspect categories is done in two steps: the first step classifies aspect terms (explicit targets), which have been detected by the CRF model presented in section 2.2 into one or more aspect category; then a second classification step is applied to classify sentences into aspect categories, to cover the cases of implicit targets (i.e. “NULL” targets).

(1) Aspect term classification into aspect categories: to achieve this task we used a precise extraction of features that are relevant for a given term in a given sentence, knowing that several terms can be present in the same sentence. We apply a term centric feature extraction, i.e. for a given term, features are: lexical semantic features associated to the term by the parser (FOOD, SERVICE ); bigrams and

trigrams involving the term; all syntactic dependencies (subject, object, modifier, attribute,...) involving the term. In other words, a term, i.e. a node in the dependency graph, is represented by the information captured by the arcs connecting this specific node to other nodes of the graph. The classification models presented in section 2.3.1 output the list of aspect categories together with their probabilities: we systematically associate the class of highest probability to a term detected by the CRF, and then associate additional categories whenever this probability is above a certain threshold; the threshold for these additional categories for a term was selected by cross-validation on the training corpora.

(2) Sentence classification into aspect categories: for this purpose, we used the same set of features as previously but at sentence level (i.e. not restricted to a given term). The classification models associate the potential sentence level aspect categories together with their probabilities; we annotate at sentence level (i.e. NULL annotation) if and only if the probability is above a given threshold, also calculated by cross-validation on the training corpora.

### 2.3.3 Polarity Classification

Opinion has to be classified with the three following polarities: positive, negative or neutral. We applied a similar strategy as for the aspect categories classification, i.e. classify the detected terms using a term-centric feature representation and then classify the sentences. We use the same pipeline as previously but in this case, we associate the highest polarity probability to the term or the sentence, ignoring the few cases presenting a mixed polarity (i.e. both positive and negative). Features are extracted the same way, but we add the aspect category detected previously as a feature for polarity classification. We also delexicalized the features, replacing a term by its generic aspect category (e.g. “staff” is replaced by “SERVICE”, “sushi” is replaced by “FOOD”, etc.), since our parser associates lexical semantic information to the domain terms.

## 3 Evaluation

Tables 1 and 2 report the results obtained on the 4 slots for English and French.

Our system performs target term classification, using a term centric representation, which proba-

	XRCE	Baseline	Rank
S1: Categories (F-1)	68.701	59.928	7/29
S2: Targets (F-1)	61.98	44.071	10/18
S12: Tuples (F-1)	48.891	37.795	2/14
S3: Polarity (Acc.)	<b>88.126</b>	76.484	1/28

**Table 1:** Results on English for the restaurant domain

	XRCE	Baseline	Rank
S1: Categories (F-1)	<b>61.207</b>	52.609	1/5
S2: Targets (F-1)	65.316	45.455	2/2
S12: Tuples (F-1)	47.721	33.017	1/1
S3: Polarity (Acc.)	<b>78.826</b>	67.4	1/5

**Table 2:** Results on French for the restaurant domain

bly explain the relatively good results we get on slot 12, <categories,target> tuples. It also explains the less good results on slot 1 for English, for which a sentence level classification would probably have been a better strategy. Results of the CRF on slot 2 are somehow disappointing, but analyzing the errors shows that we may have used a too large feature window (-3, +3) that could have resulted to an overfitting behaviour. Reducing the window to (-2, +2) improves the results: for slot 2, the F-1 reaches 67.883 for English and 68.533 for French; for slot 12, the F-1 reaches 53.364 for English, and 49.182 for French.

There are significant differences in term of performances between the two languages: for aspect category detection, the lower results for French are partly due to the smaller coverage of the lexical semantic information in the lexicon. For aspect polarity, most of the errors we’ve detected on English and French are misclassification of neutral utterances. This is due to the limited proportion of neutral cases in the training corpus (4% in English and 6% in French), and also to the fact that our polarity lexicons focus primarily on positive and negative vocabulary. This has a greater impact on the results for French, since the test corpus has a more balanced repartition of polarities.

In conclusion, the most encouraging result is that the system ranked first on polarity detection both for French and English. It tends to show that the combination of term-oriented and sentence-oriented classification performs well for polarity inference.

## 4 Conclusion

In this paper, we present a composite method based on ensemble modelling combined with rich linguistic features including lexical semantic information and syntactico-semantic dependencies to address aspect based category and polarity classification. We have also designed a target term recognizer using CRFs. Classification is performed at two levels : term level, for which we extract a set of term-centric features and sentence level, for which we extract sentence-based features, to address the cases where there is no explicit mention of a term (i.e. “NULL”). We have participated to the SemEval 2016 ABSA subtask 1, for English and French, in the restaurant domain, on slots 1, 2, 12, and 3. The system obtained very satisfying results for category detection in French (slot 1), and for slot 12 in English. But the best performances are achieved on polarity detection, since the system ends up first for both languages on slot 3: first among 28 submissions for English, and first among 5 submissions for French. Further directions of investigation will focus on two aspects. On one hand, we plan to investigate methods to decrease the level of supervision of the system (Broß, 2013), and on the other hand, we plan to extend to other languages and domains, via domain adaptation methods.

## Acknowledgements

We would like to warmly thank the SemEval 2016 Task 5 organizers and the French dataset support team for their support.

## References

- Salah Ait-Mokhtar, Jean-Pierre Chanod, and Claude Roux. 2002. Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering*, 8(2-3):121–144.
- Jurgen Broß. 2013. *Aspect-Oriented Sentiment Analysis of Customer Reviews Using Distant Supervision Techniques*. Ph.D. thesis, Freie Universität Berlin, Berlin, Germany, 7.
- Caroline Brun, Diana Nicoleta Popa, and Claude Roux. 2014. Xrce: Hybrid classification for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 838–842, Dublin, Ireland, August. Asso-

- ciation for Computational Linguistics and Dublin City University.
- Caroline Brun. 2011. Detecting opinions using deep syntactic analysis. In Galia Angelova, Kalina Bontcheva, Ruslan Mitkov, and Nicolas Nicolov, editors, *RANLP*, pages 392–398. RANLP 2011.
- Tomáš Brychcín, Michal Konkol, and Josef Steinberger. 2014. Uwb: Machine learning approach to aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 817–822, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- G. Ganu, N. Elhadad, and A. Marian. 2009. Beyond the stars: Improving rating predictions using review text content. In *Proceedings of the 12th International Workshop on the Web and Databases*, Providence, Rhode Island.
- Hussam Hamdan, Patrice Bellot, and Frederic Bechet. 2015. Lsislif: Crf and logistic regression for opinion target extraction and sentiment polarity analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 753–758, Denver, Colorado, June. Association for Computational Linguistics.
- Per C Hansen. 1986. The truncated svd as a method for regularization. Technical report, Stanford University, Stanford, CA, USA.
- Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 437–442, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- Naoaki Okazaki. 2007. Crfsuite: a fast implementation of conditional random fields (crfs).
- Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *International Workshop on Semantic Evaluation (SemEval)*.
- 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)*, pages 486–495, Denver, Colorado, June. Association for Computational Linguistics.
- Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeny Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation, SemEval '16*, San Diego, California, June. Association for Computational Linguistics.
- José Saias. 2015. Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 767–771, Denver, Colorado, June. Association for Computational Linguistics.
- Zhiqiang Toh and Wenting Wang. 2014. Dlirec: Aspect term extraction and term polarity classification system. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 235–240, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. 2014. Dcu: Aspect-based polarity classification for semeval task 4. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 392–397, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.
- H. Zou and T. Hastie. 2003. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.