Detection of Negation Cues in Spanish: The CLiC-Neg System

Javier Beltrán[1,3] and Mónica González[2,3]

[1] javier.beltran@ub.edu
[2] monica.gonzalez.manzano@gmail.com
[3] Universitat de Barcelona, Spain

Abstract. This paper describes the system proposed by the CLiC team at the University of Barcelona for detecting negation cues in Spanish. It applies the Conditional Random Field (CRF), a supervised machine learning method, to the identification of negative expressions. After carrying out an error analysis, we tried to improve on the results by the CRF adding vocabulary lists and rules. The results obtained show that, contrary to our expectations, neither adding rules nor creating a lexicon of multiword expressions significantly improved the performance of the model.

Keywords: Negation cues · Negation detection · CRF · Supervised machine learning

Resumen. Este artículo describe el sistema de detección de claves de negación propuesto por el equipo CLiC de la Universitat de Barcelona. Este sistema utiliza el Conditional Random Field (CRF), un método de aprendizaje automático supervisado, para marcar las expresiones negativas. Tras un análisis pormenorizado de los errores, hemos intentado mejorar los resultados añadiendo listas de vocabulario y reglas. Los resultados finales demuestran que, en contra de nuestras expectativas, añadir reglas o diccionarios de expresiones a un modelo de aprendizaje automático no mejora notablemente la detección de la negación.

1 Introduction

Detecting negation is a complex but pressing issue in NLP given the importance of correctly processing whether a statement is negated or not ([19]). This is especially true for tasks like sentiment analysis and areas like biomedicine, in which detecting medical conditions is crucial for a diagnosis. Negation is also a central issue for many other tasks such as information extraction and text mining.

One of the first and most influential algorithms for negation detection is NegEx ([3]), which was developed to identify negation cues in discharge summaries in English. NegEx is a rule-based algorithm and has been adapted to other languages (see [2]; [6] for German and [5] for Spanish). It has been the starting point for other extended algorithms such as ConText ([10]), pyConTextNLP or DEEPEN ([16]).

However, in recent years the research community has switched to machine learning techniques, given that this kind of approach has produced good results in other NLP related tasks. Morante & Daelemans ([17]) applied machine learning to detect negation cues and scope, Goldin & Chapman ([9]) to detect whether a term is negated and Morante et al. ([18]) specifically to detect negation scope. All these models have shown promising results and have motivated the research community to abandon rule-based algorithms in favour of machine-learning techniques. In the case of Spanish, the two best classified models in Task 2 at the NEGES 2018 Workshop [12] applied these kind of systems. Loharja et al. ([15]) used Conditional Random Fields (CRF) and Fabregat et al. ([7]) recurrent neural networks (RNN). We chose to apply CRF, following [15], with the intention of tackling one of its biggest shortcomings: the detection of discontinuous negation cues. In order to do so, we compared the results obtained by a CRF model with the results obtained by adding rules and dictionaries as a previous step to CRF classification. We think that our approach is useful for understanding more fully the limits of machine learning and demonstrating whether combining rules with this methodology actually improves results.

This paper describes our approach to solving subtask A at the NEGES 2019 conference ([11]). The goal of this subtask is to automatically identify all the negation cues in a document, including discontinuous negative expressions which, as we will show in the following sections, is the most challenging aspect in this task. The organizers provide SFU ReviewSP-NEG [13] as a corpus from which teams can build their systems. This corpus consists of 400 reviews in 8 different topics, with 25 positive and 25 negative reviews for each topic. The corpus is divided into training, development and test sets and is presented in the CoNLL format ([8]), in which each line contains a token and a column that includes grammatical information: the lemma, morphological features, and the annotation regarding negation in the case of the training and development sets.

This paper is organized as follows: Section 2 describes the different models evaluated, all of which are based on a CRF classifier and compared with a baseline that uses only a dictionary of words. Section 3 discusses the results obtained, and Section 4 synthesizes our conclusions and defines lines for future research.

2 The CLiC-Neg System

We developed several methods for predicting negation markers from the data provided. The corpus was already lemmatized and morphologically tagged, and the negation markers were indicated in the training and development sets. We translated these to a representation in IOB format ([21]), which distinguishes between the beginning and the rest of the cue. Specifically, the initial item in a complex negation cue or a simple negative expression is identified with the tag "B" (Begin), the remaining elements in the cue are identified with the tag "I" (Inside) and all the tokens that do not correspond to a negative expression are tagged as "O" (Outside). This is necessary to predict the cues correctly, as it allows one to differentiate several negation cues in the same sentence. Without the distinction between Begin and Inside, there is no way to say where one cue ends and another starts. The corpus only specifies which negation cues exist in the sentences, but they can be translated directly to IOB labels.

2.1 Baseline System

The first step in developing our model was to design a simple baseline to compare it with a more sophisticated approach. Therefore, we designed a baseline that only used a dictionary to detect negation cues. We obtained it from the SFU-ReviewSP-NEG corpus provided for the task. We first built a classifier using logistic regression to detect whether a sentence contains negation. This is a simpler problem for which the negation cues are accurate predictors. This system performed successfully at identifying negative sentences and, using only the lemma and part-of-speech of words as features, obtained an accuracy of 97%.

We took the 12 most predictive features of the logistic model, that is, the 12 words with the highest coefficient, and kept them in a dictionary to be used in our baseline. This model only detects simple negation cues, that is, it only used the tags B and O. Therefore, it was not capable of detecting discontinuous negation cues. However, even with all these limitations, this model obtained an F1 score of 73.07% (see Table 1).

2.2 CRF System

The next step was to choose a model to improve our baseline results. In line with [15], we developed a system based on a Conditional Random Field (CRF) model. This approach was the most successful in the NEGES 2018 Workshop. CRF is a method used for structured prediction, allowing one to predict sequences of labels efficiently by taking into account the labels of the previous and following words in a sentence. CRF requires less training data to obtain good results than other methods for structured predictions (see [14]), which made it a suitable choice for our corpus. Agarwal & Yu ([1]) showed that CRF is very accurate and achieves high F1 scores when detecting negation cues and the scope of negation in clinical notes, and Loharja et al. ([15]) applied it to the detection of negation cues in Spanish.

Table 1. Scores by the baseline system on the development set.

Domain	Precision	Recall	F1
Cars	70.27	55.32	61.91
Hotels	85.71	59.02	69.9
Washing Machines	90.62	64.44	75.32
Books	82.93	70.83	76.4
Mobile Phones	89.13	75.23	81.59
Music	72	69.23	70.59
Computers	80.49	63.46	70.97
Films	89.41	69.09	77.95
Average	82.57	65.82	73.07

We implemented our system using Python 3. The CRF model uses the CRF Suite library (see [15]), which is available as an extension of the Scikit-Learn library for machine learning. For the training process, our CRF model used a set of features based on the methodology in [15]: each word feature in a sentence consists of the word and its neighboring words (up to 6 positions before and up to 1 position after), and the same for the POS and its neighboring POS. The model was trained on the training part of the SFU ReviewSP-NEG corpus, using the default parameters of the CRF Suite algorithm. It was evaluated on the development part and it obtained an average F1 score of 84.18% (see Table 2, where this model is labeled **CRF**).

Our CRF system improved the baseline in all of the topics. Note that this improvement, though, was more significant in some areas, such as cars, where results were almost 12% better than the baseline, but the improvement was always around 10%. However, some topics, such as mobile phones (90.29%) or books (85.2%) obtained much better results than areas such as cars (75.86%). It would be interesting, for future research, to look up the specificities of each area to find linguistic differences that could justify these results.

2.3 CRF System with rules and list of words

We observed the false positives and false negatives yielded by the CRF model described in Section 2.2 when evaluated on the development set, in an attempt to correct the main sources of error. We came to the conclusion that most errors could be classified in four categories:

1. Errors that were caused because non-negative sentences were detected as negative: ‘Ya estaba casi todo, no (B)?’
2. Errors contained in multiword expressions: some multiword expressions that include negation cues were not correctly identified by the system (‘a_no_ser_que’, ‘a_excepción_de’, ‘a_falta_de’, etc.).
3. Errors in tagging elements such as ‘tan’, ‘tanto’ and ‘muy’, ‘mucho’, especially in the case of discontinuous cues.

4. Errors in detecting discontinuous negation cues: one particularity of negation in Spanish is that we can find more than one negation cue as part of the same negative statement.

In order to improve the CRF model, we applied two different approaches: 1) we introduced a list of rules that could help to identify multiword expressions and discontinuous cues, a common cause of misclassifications; 2) we added a dictionary list of negative expressions to reduce the false negatives that are due to unseen negation markers in the training dataset.

Rule-based system and CRF First, we applied a set of rules for correctly assigning a tag to cases where we observed that the CRF of the previous section fails. These rules are useful for detecting not only negation cues but also prevent some words from being erroneously tagged as negation. The rules are the following:

- Rule 1: In the sequence ", no?", nothing is marked as a negation cue; this is a case of not negative sentence even though it uses "no", a word generally marked as a negation cue in the CRF model in Section 2.2.
- Rule 2: If "no" is followed by "nada más" in a distance between 0 and 5 words, neither "no" nor "nada más" are negation markers.
- Rule 3: If "ningún" appears in the initial position of a sentence, it is tagged as "B".
- Rule 4: If "no", "tampoco" or "sin" are followed by "nada", "ningún" or "nadie" in a distance between 0 and 10 words, the former word is tagged as "B" and the latter as "I", thereby being a discontinuous cue.
- Rule 5: if "aún" or "todavía" are immediately followed by "no", the former is tagged as "B" and "no" is tagged as "I".
- Rule 6: "tan" or "tanto" are always tagged as "O". This avoids a common false positive in the CRF model in Section 2.2.

Then, we applied the same CRF model presented in Section 2.2 to tag the remaining words, those not classified by the rules. That is, the rules' decisions prevail over the CRF model in the cases where they are triggered. Results obtained on the development set for this model are presented in Table 2 under the name **Rules+CRF**.

List of words and CRF Our second approach uses a list of multiword expressions that were extracted from NewsCom, a corpus developed and annotated by CLiC at the University of Barcelona containing users comments on news. If a word in a sentence appears in the list, it is tagged as a negation marker. Then, we used the same CRF model to tag the remaining words, which were not detected using the list. This means that the prediction made by the CRF model only takes into account cases outside the list.

When fine-tuning this model, we observed that some words in the list also caused false positives, because they were used in other senses than negation. Similarly, we increased the list with cases of negative expressions not learned by the

CRF, such as 'a_no_ser','en_absoluto', 'en_ningún_momento' and 'sin_necesidad_de'. In Table 2 we present the evaluation results for the list of words that yielded the best score when tested on the development set, under the label **List+CRF**.

3 Discussion

Table 2 summarizes the results of our three different approaches (CRF, Rules + CRF, List + CRF) tested on the development set. Contrary to our expectations, adding rules worsened the model by 3 points. This means that our rules introduced error because they were triggered in more cases than they should have been, misclassificating some examples. This contradicts the hypothesis that combining a rule-based method with machine learning techniques can improve the model by helping it to detect cues that are not correctly identified. Similarly, adding a lexicon of multiword expressions does not improve the scores either, although the effect is less significant. Despite implementing these approaches in a controlled way, through the manual observation of the main sources of error in the predictions, we increased the number of errors rather than reduced them.

Table 2. F1 scores by the different systems tried on the development set.

Domain	CRF	List + CRF	Rules + CRF
Cars	75.86	75	76
Hotels	85.18	89.29	81
Washing Machines	86.42	86.36	86.42
Books	85.2	83.57	84.13
Mobile Phones	90.29	89.42	84.69
Music	81.48	80	77
Computers	81.25	81.25	80.85
Films	87.81	88.46	85
Average	84.18	84.17	81.89

Our best model (CRF) was presented in the NEGES 2019 competition ([11]), where it was evaluated on their test set, achieving the results shown in Table 3. It was ranked first among its competitors, but did not improve on the best score obtained at the same competition in the previous year by Loharja et al. ([15]), whose F1 score is 2% higher than ours. The main difference is in the feature set used, which is larger in their case and represents more linguistic phenomena. For instance, the use of features at the sub-word level such as prefixes or suffixes could help detect cases of morphological negation that are missing from our model.

Table 3. Scores by the chosen system (List+CRF) on the test set.

Domain	Precision	Recall	F1
Cars	94.92	82.35	88.19
Hotels	87.5	71.19	78.51
Washing Machines	92.98	76.81	84.13
Books	80.59	75.79	78.12
Mobile Phones	87.76	75.44	81.13
Music	94.44	78.16	85.53
Computers	90.48	93.83	92.12
Films	88.67	81.6	84.99
Average	89.67	79.40	84.09

4 Conclusions

In this paper we have described our proposal for the detection of negation cues based on a CRF classifier, following the approach in [15] for NEGES 2018. The results obtained show that this supervised learning technique is a promising approach to building a system that automatically detects negation cues in Spanish, with an average F1 score of 84.09% on unseen data. However, our expectations of significantly improving on these results by identifying the main sources of error and readjusting the model have not produced the desired outcome. In fact, adding rules designed to assign the right tags to some cues worsened the F1 score, and using a dictionary of negative expressions also failed to improve our results.

For future research, we will analyze the linguistic characteristics of each field to refine our CRF model. We have seen that there are significant differences in the precision and recall depending on the topic of the comments. It will be of interest to analyze whether textual differences, such as register or syntax, have a meaningful impact on the quality of our model. Additionally, the use of features at the sub-word level seems a promising approach to the detection of cases of morphological negation and the reduction of the problems related to learning from a training corpus with a limited set of negative expressions.

Acknowledgements

This work has been carried out in the framework of the MISMIS-Language project (PGC2018-096212-B-C33), funded by the Ministerio de Ciencia, Innovación y Universidades, and CLiC (2017SGR341), funded by the Agéncia de Gestió d'Ajuts Universitaris i de Recerca. Generalitat de Catalunya (AGAUR).

References

1. Agarwal, Shashank, & Hong Yu: Biomedical negation scope detection with conditional random fields. Journal of the American Medical Informatics Association, 17(6), 696–701 (2010)
2. Chapman, Wendy W., et al.: Extending the NegEx lexicon for multiple languages. Studies in health technology and informatics, 192, 677 (2013)
3. Chapman, Wendy W., et al.: A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of biomedical informatics, 34(5), 301–310 (2001)
4. Chen, Edwin: Introduction to Conditional Random Fields. http://blog.echen.me/2012/01/03/introduction-to-conditional-random-fields/. Last accessed 20 June 2019
5. Costumero, Roberto, et al.: An approach to detect negation on medical documents in Spanish. International Conference on Brain Informatics and Health. Springer, Cham, 366–375 (2014)
6. Cotik, Viviana, et al.: Negation detection in clinical reports written in German. Proceedings of the Fifth Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM2016), 115–124 (2016)
7. Fabregat, H. et al.: Deep Learning Approach for Negation Cues Detection in Spanish. Proceedings of NEGES 2018. Workshop on Negation in Spanish, 43–48 (2018)
8. Farkas, Richrd, et al.: The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. Proceedings of the Fourteenth Conference on Computational Natural Language Learning, Shared Task, Association for Computational Linguistics (2010)
9. Goldin, I., & Wendy W. Chapman: Learning to detect negation with notin medical texts. Proc Workshop on Text Analysis and Search for Bioinformatics, ACM SIGIR (2003)
10. Harkema, Henk, et al.: ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. Journal of biomedical informatics, 42(5), 839–851 (2009)
11. Jiménez-Zafra, Salud María and Cruz Díaz, Noa P. and Morante, Roser and Martín-Valdivia, María Teresa: NEGES 2019 Task: Negation in Spanish. Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2019), CEUR Workshop Proceedings, Bilbao, Spain, CEUR-WS (2019)
12. Jiménez-Zafra, Salud María and Díaz, Noa P Cruz and Morante, Roser and Martín-Valdivia, María Teresa: NEGES 2018: Workshop on Negation in Spanish. Procesamiento del Lenguaje Natural, 62, 21–28 (2019)
13. Jiménez-Zafra, S. M., Taulé, M., Martín Valdivia, M.T., Urea-Lpez, L.A., & Martí, M.A.: SFU Review SP-NEG: a Spanish-corpus annotated with negation for sentiment analysis. A typology of negation patterns. Language Resources and Evaluation, 52(2), 533–569 (2018)
14. Lafferty, John, Andrew McCallum & Fernando CN Pereira: Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001), 282–289 (2001)
15. Loharja, Henry, Lluís Padró & Jorge Turmo Borras: Negation cues detection using CRF on Spanish product review texts. NEGES 2018: Workshop on Negation in Spanish: Seville, Spain: September 19-21, 2018: proceedings book, 49–54 (2018)

16. Mehrabi, Saeed, et al.: DEEPEN: A negation detection system for clinical text incorporating dependency relation into NegEx. Journal of biomedical informatics, 54, 213–219 (2015)
17. Morante, Roser & Walter Daelemans: A metalearning approach to processing the scope of negation. Proceedings of the Thirteenth Conference on Computational Natural Language Learning, Association for Computational Linguistics, 21–29 (2009)
18. Morante, Roser, Sarah Schrauwen & Walter Daelemans: Corpus-based approaches to processing the scope of negation cues: an evaluation of the state of the art. Proceedings of the Ninth International Conference on Computational Semantics, Association for Computational Linguistics, 350–354 (2011)
19. Morante, Roser & Caroline Sporleder: Modality and Negation: An Introduction to the Special Issue. Computational Linguistics, 38(2), 223–260 (2012)
20. Prateek, Joshi: Why Do We Need Conditional Random Fields?. https://prateekvjoshi.com/2013/02/23/why-do-we-need-conditional-random-fields/. Last accessed 20 June 2019
21. Ramshaw, Lance A., & Mitchell P. Marcus: Text chunking using transformation-based learning. Natural language processing using very large corpora, Springer, Dordrecht, 157–176 (1999)