

UCM-2: a Rule-Based Approach to Infer the Scope of Negation via Dependency Parsing

Miguel Ballesteros, Alberto Díaz, Virginia Francisco,
Pablo Gervás, Jorge Carrillo de Albornoz and Laura Plaza

Natural Interaction Based on Language Group
Complutense University of Madrid
Spain

{miballes, albertodiaz, virginia}@fdi.ucm.es,
pgervas@sip.ucm.es, {jcalbornoz, lplazam}@fdi.ucm.es

Abstract

UCM-2 infers the words that are affected by negations by browsing dependency syntactic structures. It first makes use of an algorithm that detects negation cues, like *no*, *not* or *nothing*, and the words affected by them by traversing Minipar dependency structures. Second, the scope of these negation cues is computed by using a post-processing rule-based approach that takes into account the information provided by the first algorithm and simple linguistic clause boundaries. An initial version of the system was developed to handle the annotations of the Bioscope corpus. For the present version, we have changed, omitted or extended the rules and the lexicon of cues (allowing prefix and suffix negation cues, such as *impossible* or *meaningless*), to make it suitable for the present task.

1 Introduction

One of the challenges of the *SEM Shared Task (Morante and Blanco, 2012) is to infer and classify the scope and event associated to negations, given a training and a development corpus based on Conan Doyle stories (Morante and Daelemans, 2012). Negation, simple in concept, is a complex but essential phenomenon in any language. It turns an affirmative statement into a negative one, changing the meaning completely. We believe therefore that being able to handle and classify negations we would be able to improve several text mining applications.

Previous to this Shared Task, we can find several systems that handle the scope of negation in the state

of the art. This is a complex problem, because it requires, first, to find and capture the negation cues, and second, based on either syntactic or semantic representations, to identify the words that are directly (or indirectly) affected by these negation cues. One of the main works that started this trend in natural language processing was published by Morante's team (2008; 2009), in which they presented a machine learning approach for the biomedical domain evaluating it on the Bioscope corpus.

In 2010, a Workshop on Negation and Speculation in Natural Language Processing (Morante and Sporleder, 2010) was held in Uppsala, Sweden. Most of the approaches presented worked in the biomedical domain, which is the most studied in negation detection.

The system presented in this paper is a modification of the one published in Ballesteros et al. (2012). This system was developed in order to replicate (as far as possible) the annotations given in the Bioscope corpus (Vincze et al., 2008). Therefore, for the one presented in the task we needed to modify most of the rules to make it able to handle the more complex negation structures in the Conan Doyle corpus and the new challenges that it represents. The present paper has the intention of exemplifying the problems of such a system when the task is changed.

Our system presented to the Shared Task is based on the following properties: it makes use of an algorithm that traverses dependency structures, it classifies the scope of the negations by using a rule-based approach that studies linguistic clause boundaries and the outcomes of the algorithm for traversing dependency structures, it applies naive and simple

solutions to the problem of classifying the negated event and it does not use the syntactic annotation provided in the Conan Doyle corpus (just in an exception for the negated event annotation).

In Section 2 we describe the algorithms that we propose for inferring the scope of negation and the modifications that we needed to make to the previous version. In Section 3 we discuss the evaluation performed with the blind test set and development set and the error analysis over the development set. Finally, in Section 4 we give our conclusions and suggestions for future work.

2 Methodology

Our system consists of two algorithms: the first one is capable of inferring words affected by the negative operators (cues) by traversing dependency trees and the second one is capable of annotating sentences within the scope of negations. This second algorithm is the one in which we change the behaviour in a deeper way. The first one just serves as a consulting point in some of the rules of the second one. By using the training set and development set provided to the authors we modified, omitted or changed the old rules when necessary.

The first algorithm which traverses a dependency tree searching for negation cues to determine the words affected by negations, was firstly applied (at an earlier stage) to a very different domain (Ballesteros et al., 2010) obtaining interesting results. At that time, the Minipar parser (Lin, 1998) was selected to solve the problem in a simple way without needing to carry out several machine learning optimizations which are well known to be daunting tasks. We also selected Minipar because at that moment we only needed unlabelled parsing.

Therefore, our system consists of three different modules: a static negation cue lexicon, an algorithm that from a parse given by Minipar and the negation cue lexicon produces a set of words affected by the negations, and a rule-based system that produces the annotation of the scope of the studied sentence. These components are described in the following sections.

In order to annotate the sentence as it is done in the Conan Doyle corpus, we also developed a post-processing system that makes use of the outcomes

of the initial system and produces the expected output. Besides this, we also generate a very naive rule-based approach to handle the problem of annotating the negated event.

It is worth to mention that we did not make use of the syntactic annotation provided in the Conan Doyle corpus, our input is the plain text sentence. Therefore, the system could work without the columns that are included in the annotation, just with the word forms. We only make use of the annotation when we annotate the negated event, checking the part-of-speech tag to ascertain whether the corresponding word is a verb or not. The system could work without these columns but only the results of the negated event would be affected.

2.1 Negation Cue Lexicon

The lexicon containing the negation cues is static. It can be extended indefinitely but it has the restriction that it does not learn and it does not grow automatically when applying it to a different domain. The lexicon used in the previous system (Ballesteros et al., 2012) was also static but it was very small compared to the one employed by the present system, just containing less than 20 different negation cues.

Therefore, in addition to the previous lexicon, we analysed the training set and development sets and extracted 153 different negation cues (plus the ones already present in the previous system). We stored these cues in a file that feeds the system when it starts. Table 1 shows a small excerpt of the lexicon.

not	no	neither..nor
unnecessary	unoccupied	unpleasant
unpractical	unsafe	unseen
unshaven	windless	without

Table 1: Excerpt of the lexicon

2.2 Affected Wordforms Detection Algorithm

The algorithm that uses the outcomes of Minipar is the same employed in (Ballesteros et al., 2012) without modifications. It basically traverses the dependency structures and returns for each negation cue a set of words affected by the cue.

The algorithm takes into account the way of handling main verbs by Minipar, in which these verbs

appear as heads and the auxiliary verbs are dependants of them. Therefore, the system first detects the nodes that contain a word which is a negation cue, and afterwards it does the following:

- If the negation cue is a verb, such as *lack*, it is marked as a negation cue.
- If the negation cue is not a verb, the algorithm marks the main verb (if it exists) that governs the structure as a negation cue.

For the rest of nodes, if a node depends directly on any of the ones previously marked as negation cue, the system marks it as affected. The negation is also propagated until finding leaves, so wordforms that are not directly related to the cues are detected too.

Finally, by using all of the above, the algorithm generates a list of words affected by each negation cue.

2.3 Scope Classification Algorithm

This second algorithm is the one that has suffered the deepest modifications from the first version. The previous version handled the annotation as it is done in the Bioscope corpus. The algorithm works as follows:

- The system opens a scope when it finds a new negation cue detected by the affected wordforms detection algorithm. In Bioscope, only the sentences in passive voice include the subject inside the scope. However, the Conan Doyle corpus does not contain this exception always including the subject in the scope when it exists. Therefore, we modified the decision that fires this rule, and we apply the way of annotating sentences in passive voice for all the negation cues, either passive or active voice sentences.

Therefore, for most of the negation cues the system goes backward and opens the scope when it finds the subject involved or a marker that indicates another statement, like a comma.

There are some exceptions to this, such as scopes in which the cue is *without* or *neither...nor*. For them the system just opens the scope at the cue.

- The system closes a scope when there are no more wordforms to be added, i.e.:
 - It finds words that indicate another statement, such as *but* or *because*.
 - No more words in the output of the first algorithm.
 - End of the sentence.
- We also added a new rule that can handle the negation cues that are prefix or suffix of another word, such as *meaning-less*: if the system finds a cue word like this, it then annotates the suffix or prefix as the cue (such as *less*) and the rest of the word as part of the scope. Note that the Affected Wordforms Detection algorithm detects the whole word as a cue word.

2.4 Negated Event Handling

In order to come up with a solution that could provide at least some results in the negated event handling, we decided to do the following:

- When the cue word contains a negative prefix or a negative suffix, we annotate the word as the negated event.
- When the cue word is either *not* or *n't* and the next word is a verb, according to the part-of-speech annotation of the Conan Doyle corpus, we annotate the verb as the negated event.

2.5 Post-Processing Step

The post-processing step basically processes the annotated sentence with Bioscope style, (we show an example for clarification: `<scope>There is <cue>no</cue> problem</scope>`). It tokenizes the sentences, in which each token is a word or a wordform, after that, it does the following:

- If the token contains the string `<scope>`, the system just starts a new scope column reserving three new columns and it puts the word in the first free “scope” column. Because it means that there is a new scope for the present sentence.
- If the token is between a `<cue>` annotation, the system puts it in the corresponding free “cue” column of the scope already opened.

- If the token is annotated as “negated event”, the system just puts the word in the last column of the scope already opened.

Note that these three rules are not exclusive and can be fired for the same token, but in this case they are fired in the same order as they are presented.

3 Results and Discussion

In this section we first show the evaluation results and second the error analysis after studying the results on the development set.

3.1 Results

In this section we show the results obtained in two different tables: Table 2 shows the results of the system with the test set, Table 3 shows the results of the system with the development set.

As we can observe, the results for the development set are higher than the ones obtained for the test set. The reason is simple, we used the development set (apart from the training set) to modify the rules and to make the system able to annotate the sentences of the test set.

Note that our system only detects some of the negation cues (around 72% F1 and 76% F1, respectively, for the test and development sets). We therefore believe that one of the main drawbacks of the present system is the static lexicon of cues. In the previous version, due to the simplicity of the task, this was not an issue. However, it is worth noting that once the negation is detected the results are not that bad, we show a high precision in most of the tasks. But the recall suffers due to the coverage of the lexicon.

It is also worth noting that for the measure *Scope tokens*, which takes into account the tokens included in the scope but not a full scope match, our system provides interesting outcomes (around 63% F1 and 73% F1, respectively), showing that it is able to annotate the tokens in a similar way. We believe that this fact evidences that the present system comes from a different kind of annotation and a different domain, and the extension or modification of such a system is a complex task.

We can also observe that the *negated events* results are very low (around 17.46% F1 and 22.53% F1, respectively), but this was expected because by

using our two rules we are only covering two cases and moreover, these two cases are not always behaving in the same way in the corpora.

3.2 Error Analysis

In this section we analyse the different errors of our system with respect to the development set. This set contains 787 sentences, of which 144 are negation sentences containing 168 scopes, 173 cues and 122 negation events.

With respect to the negation cue detection we have obtained 58 false negatives (fn) and 16 false positives (fp). These results are not directly derived from the static lexicon of cues. The main problem is related with the management of sentences with more than one scope. The majority of the errors have been produced because in some cases all the cues are assigned to all the scopes detected in the same sentence, generating fp, and in other cases the cues of the second and subsequent scopes are ignored, generating fn. The first case occurs in sentences like (1), *no* and *without* are labelled as cues in the two scopes. The second case occurs in sentences like (2), where neither the second scope nor the second cue are labelled. In sentence (3) *un* is labelled as cue two times (unbrushed, unshaven) but within the same scope, generating a fp in the first scope and a fn in the second one.

- (1) But no [one can glance at your toilet and attire without [seeing that your disturbance dates from the moment of your waking .. ’]]
- (2) [You do n’t [mean] - . [you do] n’t [mean that I am suspected] ? ”
- (3) Our client smoothed down [his] un[brushed hair] and felt [his] un[shaven chin].

We also found false negatives that occur in multi word negation cues as *by no means*, *no more* and *rather than*.

A different kind of false positives is related to modality cues, dialogue elements and special cases (Morante and Blanco, 2012). For example, *no* in (4), *not* in (5) and *save* in (6).

- (4) “ You traced him through the telegram , no [doubt]., ” said Holmes .

Test set	gold	system	tp	fp	fn	precision (%)	recall (%)	F1 (%)
Cues:	264	235	170	39	94	81.34	64.39	71.88
Scopes(cue match):	249	233	96	47	153	67.13	38.55	48.98
Scopes(no cue match):	249	233	96	48	152	66.90	38.96	49.24
Scope tokens(no cue match):	1805	2096	1222	874	583	58.30	67.70	62.65
Negated(no cue match):	173	81	36	42	134	46.15	21.18	29.03
Full negation:	264	235	29	39	235	42.65	10.98	17.46

Table 2: Test set results.

Development	gold	system	tp	fp	fn	precision (%)	recall (%)	F1 (%)
Cues:	173	161	115	16	58	87.79	66.47	75.66
Scopes(cue match):	168	160	70	17	98	80.46	41.67	54.90
Scopes(no cue match):	168	160	70	17	98	80.46	41.67	54.90
Scope tokens(no cue match):	1348	1423	1012	411	336	71.12	75.07	73.04
Negated(no cue match):	122	71	35	31	82	53.03	29.91	38.25
Full negation:	173	161	24	16	149	60.00	13.87	22.53

Table 3: Development set results.

- (5) “ All you desire is a plain statement , [is it] not ? ’.
- (6) Telegraphic inquiries ... that [Marx knew] nothing [of his customer save that he was a good payer] .

We can also find problems with affixal negations, that is, bad separation of the affix and root of the word. For example, in (7) *dissatisfied* was erroneously divided in *di-* and *ssatisfied*. Again, it is derived from the use of a static lexicon.

- (7) He said little about the case, but from that little we gathered that [he also was not dis[satisfied] at the course of events].

Finally, we could also find cases that may be due to annotation errors. For example, *incredible* is not annotated as negation cue in (8). The annotation of this cue we think is inconsistent, it appears 5 times in the training corpus, 2 times is labelled as cue, but 3 times is not. According to the context in this sentence, *incredible* means not *credible*.

- (8) “Have just had most incredible and grotesque experience.

With respect to the full scope detection, most of the problems are due again to the management of

sentences with more than one scope. We have obtained 98 fn and 17 fp. Most of the problems are related with affixal negations, as in (9), in which all the words are included in the scope, which according to the gold standard is not correct.

- (9) [Our client looked down with a rueful face at his own] un[conventional appearance].

With respect to the scope tokens detection, the results are higher, around 73% F1 in scope tokens compared to 55% in full match scopes. The reason is because our system included tokens for the majority of scopes, increasing the recall until 75% but lowering the precision due to the inclusion of more fp.

4 Conclusions and Future Work

In this paper we presented our participation in the SEM-Shared Task, with a modification of a rule-based system that was designed to be used in a different domain. As the main conclusion we could say that modifying such a system to perform in a different type of texts is complicated. However, taking into account this fact, and the results obtained, we are tempted to say that our system presents competitive results.

We believe that the present system has a lot of room for improvement: (i) improve the management of sentences with more than one scope modifying the scope classification algorithm and the post-processing step, (ii) replacing the dependency parser with a state-of-the-art parser in order to get higher performance, or (iii) proposing a different way of getting a reliable lexicon of cues, by using a semantic approach that informs if the word has a negative meaning in the context of the sentence. Again, this could be achieved by using one of the parsers presented in the ConLL 2008 Shared Task (Surdeanu et al., 2008).

Acknowledgments

This research is funded by the Spanish Ministry of Education and Science (TIN2009-14659-C03-01 Project).

References

- Miguel Ballesteros, Raúl Martín, and Belén Díaz-Agudo. 2010. Jadaweb: A cbr system for cooking recipes. In *Proceedings of the Computing Cooking Contest of the International Conference of Case-Based Reasoning*.
- Miguel Ballesteros, Virginia Francisco, Alberto Díaz, Jesús Herrera, and Pablo Gervás. 2012. Inferring the scope of negation in biomedical documents. In *Proceedings of the 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2012)*, New Delhi. Springer.
- Dekang Lin. 1998. Dependency-based evaluation of MINIPAR. In *Proceedings of the Workshop on the Evaluation of Parsing Systems*, Granada.
- Roser Morante and Eduardo Blanco. 2012. Sem 2012 shared task: Resolving the scope and focus of negation. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM 2012)*, Montreal, Canada.
- Roser Morante and Walter Daelemans. 2009. A meta-learning approach to processing the scope of negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, CoNLL '09, pages 21–29, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Roser Morante and Walter Daelemans. 2012. Conandoyle-neg: Annotation of negation in conandoyle stories. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*. Istanbul, Turkey.
- Roser Morante and Caroline Sporleder, editors. 2010. *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, Uppsala, Sweden.
- Roser Morante, Anthony Liekens, and Walter Daelemans. 2008. Learning the scope of negation in biomedical texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '08, pages 715–724, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *CoNLL 2008: Proceedings of the Twelfth Conference on Natural Language Learning*, pages 159–177, Manchester, United Kingdom.
- Veronika Vincze, Gyorgy Szarvas, Richard Farkas, Gyorgy Mora, and Janos Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11):S9+.