

Automatic Extraction of Lexico-Syntactic Patterns for Detection of Negation and Speculation Scopes

Emilia Apostolova

DePaul University
Chicago, IL USA

emilia.aposto@gmail.com

Noriko Tomuro

DePaul University
Chicago, IL USA

tomuro@cs.depaul.edu

Dina Demner-Fushman

National Library of Medicine
Bethesda, MD USA

ddemner@mail.nih.gov

Abstract

Detecting the linguistic scope of negated and speculated information in text is an important Information Extraction task. This paper presents ScopeFinder, a linguistically motivated rule-based system for the detection of negation and speculation scopes. The system rule set consists of lexico-syntactic patterns automatically extracted from a corpus annotated with negation/speculation cues and their scopes (the BioScope corpus). The system performs on par with state-of-the-art machine learning systems. Additionally, the intuitive and linguistically motivated rules will allow for manual adaptation of the rule set to new domains and corpora.

1 Motivation

Information Extraction (IE) systems often face the problem of distinguishing between affirmed, negated, and speculative information in text. For example, sentiment analysis systems need to detect negation for accurate polarity classification. Similarly, medical IE systems need to differentiate between affirmed, negated, and speculated (possible) medical conditions.

The importance of the task of negation and speculation (a.k.a. hedge) detection is attested by a number of research initiatives. The creation of the BioScope corpus (Vincze et al., 2008) assisted in the development and evaluation of several negation/hedge scope detection systems. The corpus consists of medical and biological texts annotated for negation, speculation, and their linguistic scope. The 2010

i2b2 NLP Shared Task¹ included a track for detection of the assertion status of medical problems (e.g. affirmed, negated, hypothesized, etc.). The CoNLL-2010 Shared Task (Farkas et al., 2010) focused on detecting hedges and their scopes in Wikipedia articles and biomedical texts.

In this paper, we present a linguistically motivated rule-based system for the detection of negation and speculation scopes that performs on par with state-of-the-art machine learning systems. The rules used by the ScopeFinder system are automatically extracted from the BioScope corpus and encode lexico-syntactic patterns in a user-friendly format. While the system was developed and tested using a biomedical corpus, the rule extraction mechanism is not domain-specific. In addition, the linguistically motivated rule encoding allows for manual adaptation to new domains and corpora.

2 Task Definition

Negation/Speculation detection is typically broken down into two sub-tasks - discovering a negation/speculation cue and establishing its scope. The following example from the BioScope corpus shows the annotated hedging cue (in bold) together with its associated scope (surrounded by curly brackets):

*Finally, we explored the {**possible** role of 5-hydroxyeicosatetraenoic acid as a regulator of arachidonic acid liberation}.*

Typically, systems first identify negation/speculation cues and subsequently try to identify their associated cue scope. However, the two tasks are interrelated and both require

¹<https://www.i2b2.org/NLP/Relations/>

syntactic understanding. Consider the following two sentences from the BioScope corpus:

1) *By contrast, {D-mib **appears** to be uniformly expressed in imaginal discs }.*

2) *Differentiation assays using water soluble phorbol esters reveal that differentiation becomes irreversible soon after AP-1 **appears**.*

Both sentences contain the word form *appears*, however in the first sentence the word marks a hedging cue, while in the second sentence the word does not suggest speculation.

Unlike previous work, we do not attempt to identify negation/speculation cues independently of their scopes. Instead, we concentrate on scope detection, simultaneously detecting corresponding cues.

3 Dataset

We used the BioScope corpus (Vincze et al., 2008) to develop our system and evaluate its performance. To our knowledge, the BioScope corpus is the only publicly available dataset annotated with negation/speculation cues and their scopes. It consists of biomedical papers, abstracts, and clinical reports (corpus statistics are shown in Tables 1 and 2).

Corpus Type	Sentences	Documents	Mean Document Size
Clinical	7520	1954	3.85
Full Papers	3352	9	372.44
Paper Abstracts	14565	1273	11.44

Table 1: Statistics of the BioScope corpus. Document sizes represent number of sentences.

Corpus Type	Negation Cues	Speculation Cues	Negation	Speculation
Clinical	872	1137	6.6%	13.4%
Full Papers	378	682	13.76%	22.29%
Paper Abstracts	1757	2694	13.45%	17.69%

Table 2: Statistics of the BioScope corpus. The 2nd and 3d columns show the total number of cues within the datasets; the 4th and 5th columns show the percentage of negated and speculative sentences.

70% of the corpus documents (randomly selected) were used to develop the ScopeFinder system (i.e. extract lexico-syntactic rules) and the remaining 30% were used to evaluate system performance. While the corpus focuses on the biomedical domain, our rule extraction method is not domain specific and in future work we are planning to apply our method on different types of corpora.

4 Method

Intuitively, rules for detecting both speculation and negation scopes could be concisely expressed as a

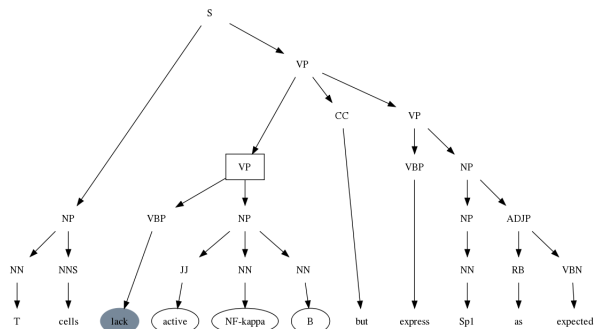


Figure 1: Parse tree of the sentence ‘*T cells {**lack** active NF-kappa B } but express Sp1 as expected*’ generated by the Stanford parser. Speculation scope words are shown in ellipsis. The cue word is shown in grey. The nearest common ancestor of all cue and scope leaf nodes is shown in a box.

combination of lexical and syntactic patterns. For example, Özgür and Radev (2009) examined sample BioScope sentences and developed hedging scope rules such as:

The scope of a modal verb cue (e.g. may, might, could) is the verb phrase to which it is attached;

The scope of a verb cue (e.g. appears, seems) followed by an infinitival clause extends to the whole sentence.

Similar lexico-syntactic rules have been also manually compiled and used in a number of hedge scope detection systems, e.g. (Kilicoglu and Bergler, 2008), (Rei and Briscoe, 2010), (Velldal et al., 2010), (Kilicoglu and Bergler, 2010), (Zhou et al., 2010).

However, manually creating a comprehensive set of such lexico-syntactic scope rules is a laborious and time-consuming process. In addition, such an approach relies heavily on the availability of accurately parsed sentences, which could be problematic for domains such as biomedical texts (Clegg and Shepherd, 2007; McClosky and Charniak, 2008).

Instead, we attempted to automatically extract lexico-syntactic scope rules from the BioScope corpus, relying only on consistent (but not necessarily accurate) parse tree representations.

We first parsed each sentence in the training dataset which contained a negation or speculation cue using the Stanford parser (Klein and Manning, 2003; De Marneffe et al., 2006). Figure 1 shows the parse tree of a sample sentence containing a negation cue and its scope.

Next, for each cue-scope instance within the sentence, we identified the nearest common ancestor

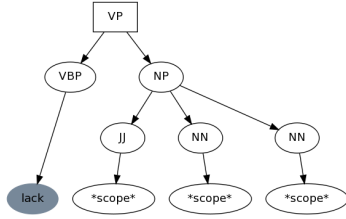
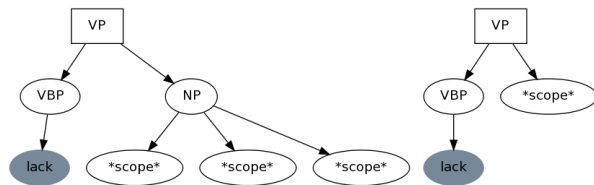


Figure 2: Lexico-syntactic pattern extracted from the sentence from Figure 1. The rule is equivalent to the following string representation: $(VP (VBP\ lack) (NP (JJ\ *scope*) (NN\ *scope*) (NN\ *scope*)))$.

which encompassed the cue word(s) and all words in the scope (shown in a box on Figure 1). The subtree rooted by this ancestor is the basis for the resulting lexico-syntactic rule. The leaf nodes of the resulting subtree were converted to a generalized representation: scope words were converted to **scope**; non-cue and non-scope words were converted to ***; cue words were converted to lower case. Figure 2 shows the resulting rule.

This rule generation approach resulted in a large number of very specific rule patterns - 1,681 negation scope rules and 3,043 speculation scope rules were extracted from the training dataset.

To identify a more general set of rules (and increase recall) we next performed a simple transformation of the derived rule set. If all children of a rule tree node are of type **scope** or *** (i.e. non-cue words), the node label is replaced by **scope** or *** respectively, and the node’s children are pruned from the rule tree; neighboring identical siblings of type **scope** or *** are replaced by a single node of the corresponding type. Figure 3 shows an example of this transformation.



(a) The children of nodes JJ/NN/NN are pruned and their labels are replaced by **scope**. (b) The children of node NP are pruned and its label is replaced by **scope**.

Figure 3: Transformation of the tree shown in Figure 2. The final rule is equivalent to the following string representation: $(VP (VBP\ lack) *scope*)$

The rule tree pruning described above reduced the negation scope rule patterns to 439 and the speculation rule patterns to 1,000.

In addition to generating a set of scope finding rules, we also implemented a module that parses string representations of the lexico-syntactic rules and performs subtree matching. The ScopeFinder module² identifies negation and speculation scopes in sentence parse trees using string-encoded lexico-syntactic patterns. Candidate sentence parse subtrees are first identified by matching the path of cue leaf nodes to the root of the rule subtree pattern. If an identical path exists in the sentence, the root of the candidate subtree is thus also identified. The candidate subtree is evaluated for a match by recursively comparing all node children (starting from the root of the subtree) to the rule pattern subtree. Nodes of type **scope** and *** match any number of nodes, similar to the semantics of Regex Kleene star (*).

5 Results

As an informed baseline, we used a previously developed rule-based system for negation and speculation scope discovery (Apostolova and Tomuro, 2010). The system, inspired by the NegEx algorithm (Chapman et al., 2001), uses a list of phrases split into subsets (preceding vs. following their scope) to identify cues using string matching. The cue scopes extend from the cue to the beginning or end of the sentence, depending on the cue type. Table 3 shows the baseline results.

Negation	Correctly Predicted Cues			All Predicted Cues
	P	R	F	F
Clinical	94.12	97.61	95.18	85.66
Full Papers	54.45	80.12	64.01	51.78
Paper Abstracts	63.04	85.13	72.31	59.86
Speculation				
Clinical	65.87	53.27	58.90	50.84
Full Papers	58.27	52.83	55.41	29.06
Paper Abstracts	73.12	64.50	68.54	38.21

Table 3: Baseline system performance. P (Precision), R (Recall), and F (F1-score) are computed based on the sentence tokens of correctly predicted cues. The last column shows the F1-score for sentence tokens of all predicted cues (including erroneous ones).

We used only the scopes of predicted cues (correctly predicted cues vs. all predicted cues) to mea-

²The rule sets and source code are publicly available at <http://scopefinder.sourceforge.net/>.

sure the baseline system performance. The baseline system heuristics did not contain all phrase cues present in the dataset. The scopes of cues that are missing from the baseline system were not included in the results. As the baseline system was not penalized for missing cue phrases, the results represent the upper bound of the system.

Table 4 shows the results from applying the full extracted rule set (1,681 negation scope rules and 3,043 speculation scope rules) on the test data. As expected, this rule set consisting of very specific scope matching rules resulted in very high precision and very low recall.

Negation	P	R	F	A
Clinical	99.47	34.30	51.01	17.58
Full Papers	95.23	25.89	40.72	28.00
Paper Abstracts	87.33	05.78	10.84	07.85
Speculation				
Clinical	96.50	20.12	33.30	22.90
Full Papers	88.72	15.89	26.95	10.13
Paper Abstracts	77.50	11.89	20.62	10.00

Table 4: Results from applying the full extracted rule set on the test data. Precision (P), Recall (R), and F1-score (F) are computed based on the number of correctly identified scope tokens in each sentence. Accuracy (A) is computed for correctly identified full scopes (exact match).

Table 5 shows the results from applying the rule set consisting of pruned pattern trees (439 negation scope rules and 1,000 speculation scope rules) on the test data. As shown, overall results improved significantly, both over the baseline and over the unpruned set of rules. Comparable results are shown in bold in Tables 3, 4, and 5.

Negation	P	R	F	A
Clinical	85.59	92.15	88.75	85.56
Full Papers	49.17	94.82	64.76	71.26
Paper Abstracts	61.48	92.64	73.91	80.63
Speculation				
Clinical	67.25	86.24	75.57	71.35
Full Papers	65.96	98.43	78.99	52.63
Paper Abstracts	60.24	95.48	73.87	65.28

Table 5: Results from applying the pruned rule set on the test data. Precision (P), Recall (R), and F1-score (F) are computed based on the number of correctly identified scope tokens in each sentence. Accuracy (A) is computed for correctly identified full scopes (exact match).

6 Related Work

Interest in the task of identifying negation and speculation scopes has developed in recent years. Rele-

vant research was facilitated by the appearance of a publicly available annotated corpus. All systems described below were developed and evaluated against the BioScope corpus (Vincze et al., 2008).

Özgür and Radev (2009) have developed a supervised classifier for identifying speculation cues and a manually compiled list of lexico-syntactic rules for identifying their scopes. For the performance of the rule based system on identifying speculation scopes, they report 61.13 and 79.89 accuracy for BioScope full papers and abstracts respectively.

Similarly, Morante and Daelemans (2009b) developed a machine learning system for identifying hedging cues and their scopes. They modeled the scope finding problem as a classification task that determines if a sentence token is the first token in a scope sequence, the last one, or neither. Results of the scope finding system with predicted hedge signals were reported as F1-scores of 38.16, 59.66, 78.54 and for clinical texts, full papers, and abstracts respectively³. Accuracy (computed for correctly identified scopes) was reported as 26.21, 35.92, and 65.55 for clinical texts, papers, and abstracts respectively.

Morante and Daelemans have also developed a metalearner for identifying the scope of negation (2009a). Results of the negation scope finding system with predicted cues are reported as F1-scores (computed on scope tokens) of 84.20, 70.94, and 82.60 for clinical texts, papers, and abstracts respectively. Accuracy (the percent of correctly identified exact scopes) is reported as 70.75, 41.00, and 66.07 for clinical texts, papers, and abstracts respectively.

The top three best performers on the CoNLL-2010 shared task on hedge scope detection (Farkas et al., 2010) report an F1-score for correctly identified hedge cues and their scopes ranging from 55.3 to 57.3. The shared task evaluation metrics used stricter matching criteria based on exact match of both cues and their corresponding scopes⁴.

CoNLL-2010 shared task participants applied a variety of rule-based and machine learning methods

³F1-scores are computed based on scope tokens. Unlike our evaluation metric, scope token matches are computed for each cue within a sentence, i.e. a token is evaluated multiple times if it belongs to more than one cue scope.

⁴Our system does not focus on individual cue-scope pair detection (we instead optimized scope detection) and as a result performance metrics are not directly comparable.

on the task - Morante et al. (2010) used a memory-based classifier based on the k-nearest neighbor rule to determine if a token is the first token in a scope sequence, the last, or neither; Rei and Briscoe (2010) used a combination of manually compiled rules, a CRF classifier, and a sequence of post-processing steps on the same task; Velldal et al (2010) manually compiled a set of heuristics based on syntactic information taken from dependency structures.

7 Discussion

We presented a method for automatic extraction of lexico-syntactic rules for negation/speculation scopes from an annotated corpus. The developed ScopeFinder system, based on the automatically extracted rule sets, was compared to a baseline rule-based system that does not use syntactic information. The ScopeFinder system outperformed the baseline system in all cases and exhibited results comparable to complex feature-based, machine-learning systems.

In future work, we will explore the use of statistically based methods for the creation of an optimum set of lexico-syntactic tree patterns and will evaluate the system performance on texts from different domains.

References

- E. Apostolova and N. Tomuro. 2010. Exploring surface-level heuristics for negation and speculation discovery in clinical texts. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pages 81–82. Association for Computational Linguistics.
- W.W. Chapman, W. Bridewell, P. Hanbury, G.F. Cooper, and B.G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*, 34(5):301–310.
- A.B. Clegg and A.J. Shepherd. 2007. Benchmarking natural-language parsers for biological applications using dependency graphs. *BMC bioinformatics*, 8(1):24.
- M.C. De Marneffe, B. MacCartney, and C.D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *LREC 2006*. Citeseer.
- R. Farkas, V. Vincze, G. Móra, J. Csirik, and G. Szarvas. 2010. The CoNLL-2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010): Shared Task*, pages 1–12.
- H. Kilicoglu and S. Bergler. 2008. Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC bioinformatics*, 9(Suppl 11):S10.
- H. Kilicoglu and S. Bergler. 2010. A High-Precision Approach to Detecting Hedges and Their Scopes. *CoNLL-2010: Shared Task*, page 70.
- D. Klein and C.D. Manning. 2003. Fast exact inference with a factored model for natural language parsing. *Advances in neural information processing systems*, pages 3–10.
- D. McClosky and E. Charniak. 2008. Self-training for biomedical parsing. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pages 101–104. Association for Computational Linguistics.
- R. Morante and W. Daelemans. 2009a. A metalearning approach to processing the scope of negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 21–29. Association for Computational Linguistics.
- R. Morante and W. Daelemans. 2009b. Learning the scope of hedge cues in biomedical texts. In *Proceedings of the Workshop on BioNLP*, pages 28–36. Association for Computational Linguistics.
- R. Morante, V. Van Asch, and W. Daelemans. 2010. Memory-based resolution of in-sentence scopes of hedge cues. *CoNLL-2010: Shared Task*, page 40.
- A. Özgür and D.R. Radev. 2009. Detecting speculations and their scopes in scientific text. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, pages 1398–1407. Association for Computational Linguistics.
- M. Rei and T. Briscoe. 2010. Combining manual rules and supervised learning for hedge cue and scope detection. In *Proceedings of the 14th Conference on Natural Language Learning*, pages 56–63.
- E. Velldal, L. Øvrelid, and S. Oepen. 2010. Resolving Speculation: MaxEnt Cue Classification and Dependency-Based Scope Rules. *CoNLL-2010: Shared Task*, page 48.
- V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC bioinformatics*, 9(Suppl 11):S9.
- H. Zhou, X. Li, D. Huang, Z. Li, and Y. Yang. 2010. Exploiting Multi-Features to Detect Hedges and Their Scope in Biomedical Texts. *CoNLL-2010: Shared Task*, page 106.