

UMichigan: A Conditional Random Field Model for Resolving the Scope of Negation

Amjad Abu-Jbara
EECS Department
University of Michigan
Ann Arbor, MI, USA
amjbara@umich.edu

Dragomir Radev
EECS Department
University of Michigan
Ann Arbor, MI, USA
radev@umich.edu

Abstract

In this paper, we present a system for detecting negation in English text. We address three tasks: negation cue detection, negation scope resolution and negated event identification. We pose these tasks as sequence labeling problems. For each task, we train a Conditional Random Field (CRF) model on lexical, structural, and syntactic features extracted from labeled data. The models are trained and tested using the dataset distributed with the *sem Shared Task 2012 on resolving the scope and focus of negation. The system detects negation cues with 90.98% F1 measure (94.3% and 87.88% recall). It identifies negation scope with 82.70% F1 on token-by-token level and 64.78% F1 on full scope level. Negated events are detected with 51.10% F1 measure.

1 Introduction

Negation is a linguistic phenomenon present in all languages (Tottie, 1993; Horn, 1989). The semantic function of negation is to transform an affirmative statement into its opposite meaning. The automatic detection of negation and its scope is a problem encountered in a wide range of natural language processing applications including, but not limited to, data mining, relation extraction, question answering, and sentiment analysis. For example, failing to account for negation may result in giving wrong answers in question answering systems or in the prediction of opposite sentiment in sentiment analysis systems.

The occurrence of negation in a sentence is determined by the presence of a negation cue. A negation cue is a word, a phrase, a prefix, or a postfix that triggers negation. Scope of negation is the part of the meaning that is negated (Huddleston and Pullum, 2002). The negated event is the event or the entity that the negation indicates its absence or denies its occurrence. For example, in the sentence below *never* is the negation cue. The scope is enclosed in square brackets. The negated event is underlined.

*[Andrew had] never [liked smart phones],
but he received one as a gift last week and
started to use it.*

Negation cues and scopes may be discontinuous. For example, the negation cue *neither ... nor* is discontinuous.

In this chapter, we present a system for automatically detecting negation cues, negated events, and negation scopes in English text. The system uses conditional random field (CRF) models trained on labeled sentences extracted from two classical English novels. The CRF models are trained using lexical, structural, and syntactic features. The experiments show promising results.

This paper is organized as follows. Section 2 reviews previous work. Section 3 describes the data. Section 4 describes the CRFs models. Section 5 presents evaluation, results, and discussion.

2 Previous Work

Most research on negation has been done in the biomedical domain (Chapman et al., 2001; Mutalik et al., 2001; Kim and Park, 2006; Morante et al.,

Token	Lemma	POS	Syntax	Cue 1	Scope 1	Event 1	Cue 2	Scope 2	Event 2
She	She	PRP	(S(NP*))	-	She	-	-	-	-
would	would	MD	(VP*)	-	would	-	-	-	-
not	not	RB	*	not	-	-	-	-	-
have	have	VB	(VP*)	-	have	-	-	-	-
said	say	VBD	(VP*)	-	said	-	-	-	-
,	,	“	(SBAR(S(NP*))	-	,	-	-	-	-
Godspeed	Godspeed	NNP	*	-	Godspeed	-	-	-	-
,	,	”	*)	-	,	-	-	-	-
had	have	VBD	(VP*)	-	had	-	-	had	-
it	it	PRP	(ADVP*)	-	it	-	-	it	-
not	not	RB	*)	-	not	-	not	-	-
been	be	VBN	(VP*)	-	been	-	-	been	-
so	so	RB	(ADV*)	-	so	-	-	so	-
.	.	.	*)	-	.	-	-	.	-

Table 1: Example sentence annotated for negation following sem shared task 2012 format

2008a; Morante and Daelemans, 2009; Agarwal and Yu, 2010; Morante, 2010; Read et al., 2011), mostly on clinical reports. The reason is that most NLP research in the biomedical domain is interested in automatically extracting factual relations and pieces of information from unstructured data. Negation detection is important here because information that falls in the scope of a negation cue cannot be treated as facts.

Chapman et al. (2001) proposed a rule-based algorithm called *NegEx* for determining whether a finding or disease mentioned within narrative medical reports is present or absent. The algorithm uses regular-expression-based rules. Mutalik et al. (2001) developed another rule based system called *Negfinder* that recognizes negation patterns in biomedical text. It consists of two components: a lexical scanner, *lexer* that uses regular expression rules to generate a finite state machine, and a parser. Morante (2008b) proposed a supervised approach for detecting negation cues and their scopes in biomedical text. Their system consists of two memory-based engines, one that decides if the tokens in a sentence are negation signals, and another one that finds the full scope of these negation signals.

Negation has been also studied in the context of sentiment analysis (Wilson et al., 2005; Jia et al., 2009; Councill et al., 2010; Heerschop et al., 2011; Hogenboom et al., 2011). Wiegand et al. (2010) surveyed the recent work on negation scope detection for sentiment analysis. Wilson et al. (2005) studied

the contextual features that affect text polarity. They used a machine learning approach in which negation is encoded using several features. One feature checks whether a negation expression occurs in a fixed window of four words preceding the polar expression. Another feature accounts for a polar predicate having a negated subject. They also have disambiguation features to handle negation words that do not function as negation cues in certain contexts, e.g. *not to mention* and *not just*.

Jia et al. (2009) proposed a rule based method to determine the polarity of sentiments when one or more occurrences of a negation term such as not appear in a sentence. The hand-crafted rules are applied to syntactic and dependency parse tree representations of the sentence.

Hogenboom et al. (2011) found that applying a simple rule that considers two words, following a negation keyword, to be negated by that keyword, to be effective in improving the accuracy of sentiment analysis in movie reviews. This simple method yields a significant increase in overall sentiment classification accuracy and macro-level F1 of 5.5% and 6.2%, respectively, compared to not accounting for negation.

This work is characterized by addressing three tasks at once: negation cue detection, negated event identification, and negation scope resolution. Our proposed approach uses a supervised graphical probabilistic model trained using labeled data.

3 Data

We use the dataset distributed by the organizers of the *sem Shared Task 2012 on resolving the scope and focus of negation. This dataset includes two stories by Conan Doyle, *The Hound of the Baskervilles*, *The Adventures of Wisteria Lodge*. All occurrences of negation are annotated accounting for negation expressed by nouns, pronouns, verbs, adverbs, determiners, conjunctions and prepositions. For each negation cue, the negation cue and scope are marked, as well as the negated event (if any exists). The annotation guidelines follow the proposal of Morante et al. (2011)¹. The data is split into three sets: a training set containing 3,644 sentences, a development set containing 787 sentences, and a testing set containing 1,089 sentences. The data is provided in CoNLL format. Each line corresponds to a token and each annotation is provided in a column; empty lines indicate end of sentences. The provided annotations are:

- Column 1: chapter name
- Column 2: sentence number within chapter
- Column 3: token number within sentence
- Column 4: word
- Column 5: lemma
- Column 6: part-of-speech
- Column 7: syntax
- Columns 8 to last:
 - If the sentence has no negations, column 8 has a ”****” value and there are no more columns.
 - If the sentence has negations, the annotation for each negation is provided in three columns. The first column contains the word or part of the word (e.g., morpheme ”un”), that belongs to the negation cue. The second contains the word or part of the word that belongs to the scope of the negation cue. The third column contains the word or part of the word that is the

¹<http://www.clips.ua.ac.be/sites/default/files/ctrs-n3.pdf>

Token	Lemma	Punc.	Cat.	POS	Label
Since	Since	0	OTH	IN	O
we	we	0	PRO	PRP	O
have	have	0	VB	VBP	O
been	be	0	VB	VCN	O
so	so	0	ADVB	RB	O
unfortunate	unfortunate	0	ADJ	JJ	PRE
as	as	0	ADVB	RB	O
to	to	0	OTH	TO	O
miss	miss	0	VB	VB	O
him	him	0	PRO	PRP	O
and	and	0	OTH	CC	O
have	have	0	VB	VBP	O
no	no	0	OTH	DT	NEG
notion	notion	0	NOUN	NN	O
of	of	0	OTH	IN	O
his	his	0	PRO	PRP\$	O
errand	errand	0	NOUN	NN	O
,	,	1	OTH	,	O
this	this	0	OTH	DT	O
accidental	accidental	0	ADJ	JJ	O
souvenir	souvenir	0	NOUN	NN	O
becomes	become	0	VB	VBZ	O
of	of	0	OTH	IN	O
importance	importance	0	NOUN	NN	O
.	.	1	OTH	.	O

Table 2: Example sentence labeled for negation cue detection

negated event or property. It can be the case that no negated event or property are marked as negated.

Table 1 shows an example of an annotated sentence that contains two negation cues.

4 Approach

The problem that we are trying to solve can be split into three tasks. The first task is to detect negation cues. The second task is to identify the scope of each detected negation cue. The third task is to identify the negated event. We use a machine learning approach to address these tasks. We train a Conditional Random Field (CRF) (Lafferty et al., 2001) model on lexical, structural, and syntactic features extracted from the training dataset. In the following subsections, we describe the CRF model that we use for each task.

4.1 Negation Cue Detection

Negation cues are lexical elements that indicate the existence of negation in a sentence. From lexical

point of view, negation cues can be divided into four categories:

1. Prefix (i.e. in-, un-, im-, il-, dis-). For example, **un-** in *unsuitable* is a prefix negation cue.
2. Postfix (i.e. -less). For example, **-less** in *careless*.
3. Multi-word negation cues such as *neither...nor*, *rather than*, *by no means*, etc.
4. Single word negation cues such as *not*, *no*, *none*, *nobody*, etc.

The goal of this task is to detect negation cues. We pose this problem as a sequence labeling task. The reason for this choice is that some negation cues may not indicate negation in some contexts. For example, the negation cue *not* in the phrase *not to mention* does not indicate negation. Also, as we saw above, some negation cues may consist of multiple words, some of them are continuous and others are discontinuous. Treating the task as a sequence labeling problem help model the contextual factors that affect the function of negation cues. We train a CRF model using features extracted from the sentences of the training dataset. The token level features that we train the model on are:

- *Token*: The word or the punctuation mark as it appears in the sentence.
- *Lemma*: The lemmatized form of the token.
- *Part-Of-Speech tag*: The part of speech tag of the token.
- *Part-Of-Speech tag category*: Part-of-speech tags reduced into 5 categories: Adjective (ADJ), Verb (VB), Noun (NN), Adverb (ADVB), Pronoun (PRO), and other (OTH).
- *Is punctuation mark*: This feature takes the value 1 if the token is a punctuation mark and 0 otherwise.
- *Starts with negation prefix*: This feature takes the value 1 if the token is a word that starts with un-, in-, im-, il-, or dis- and 0 otherwise.

- *Ends with negation postfix*: This feature takes the value 1 if the token is a word that ends with -less and 0 otherwise.

The CRF model that we use considers at each token the features of the current token, the two preceding tokens, and the two proceeding tokens. The model also uses token bigrams and trigrams, and part-of-speech tag bigrams and trigrams as features.

The labels are 5 types: "O" for tokens that are not part of any negation cue; "NEG" for single word negation cues; "PRE" for prefix negation cue; "POST" for postfix negation cue; and "MULTI-NEG" for multi-word negation cues. Table 2 shows an example labeled sentence.

At testing time, if a token is labeled "NEG" or "MULTI-NEG" the whole token is treated as a negation cue or part of a negation cue respectively. If a token is labeled as "PRE" or "POST", a regular expression is used to determine the prefix/postfix that triggered the negation.

4.2 Negation Scope Detection

Scope of negation is the sequence of tokens (can be discontinuous) that expresses the meaning that is meant to be negated by a negation cue. A sentence may contain zero or more negation cues. Each negation cue has its own scope. It is possible that the scope of two negation cues overlap. We use each negation instance (i.e. each negation cue and its scope) as one training example. Therefore, a sentence that contains two negation cues provides two training examples. We train a CRF model on features extracted from all negation instances in the training dataset. The features that we use are:

- *Token*: The word or the punctuation mark as it appears in the sentence.
- *Lemma*: The lemmatized form of the token.
- *Part-Of-Speech tag*: The part of speech tag of the token.
- *Part-Of-Speech tag category*: Part-of-speech tags reduced into 5 categories: Adjective (ADJ), Verb (VB), Noun (NN), Adverb (ADVB), Pronoun (PRO), and other (OTH).

- *Is punctuation mark*: This feature takes the value 1 if the token is a punctuation mark and 0 otherwise.
- *Type of negation cue*: Possible types are: "NEG" for single word negation cues; "PRE" for prefix negation cue; "POST" for postfix negation cue; and "MULTI" for multi-word negation cues.
- *Relative position*: This feature takes the value 1 if the token position in the sentence is before the position of the negation cue, 2 if the token position is after the position of the negation cue, and 3 if the token is the negation cue itself.
- *Distance*: The number of tokens between the current token and the negation cue.
- *Same segment*: This feature takes the value 1 if this token and the negation cue fall in the segment in the sentence. The sentence is segmented by punctuation marks.
- *Chunk*: This feature takes the value NP-B (VP-B) if this token is the first token of a noun (verb) phrase, NP-I (VP-I) if it is inside a noun (verb) phrase, NP-E (VP-E) if it is the last token of a noun (verb) phrase.
- *Same chunk*: This feature takes the value 1 if this token and the negation cue fall in the same chunk (noun phrase or verb phrase).
- *Is negation*: This feature takes the value 1 if this token is a negation cue, and 0 otherwise.
- *Syntactic distance*: The number of edges in the shortest path that connects the token and the negation in the syntactic parse tree.
- *Common ancestor node*: The type of the node in the syntactic parse tree that is the least common ancestor of this token and the negation cue token.

The CRF model considers the features of 4 tokens to the left and to the right at each position. It also uses bigram and trigram combinations of some of the features.

At testing time a few postprocessing rules are used to fix sure labels if they were labeled incorrectly. For example, if a word starts with a prefix negation cue, the word itself (without the prefix) is always part of the scope and it is also the negated event.

4.3 Negated Event Identification

It is possible that a negation cue comes associated with an event. A negation has an event if it occurs in a factual context. The dataset that we use was labeled for negated events whenever one exists. We used the same features described in the previous subsection to train a CRF model for negated event identification. We have also tried to use one CRF model for both scope resolution and negated event identification, but we noticed that using two separate models results in significantly better results for both tasks.

5 Evaluation

We use the testing set described in Section 3 to evaluate the system. The testing set contains 1089 sentences 235 of which contains at least one negation.

We use the standard precision, recall, and f-measure metrics to evaluate the system. We perform the evaluation on different levels:

1. Cues: the metrics are computed only for cue detection.
2. Scope (tokens): the metrics are calculated at token level. If a sentence has 2 scopes, one with 5 tokens and another with 4, the total number of scope tokens is 9.
3. Scope (full): the metrics are calculated at the full scope level. Both the negation cue and the whole scope should be correctly identified. If a sentence contains 2 negation cues, then 2 scopes are checked. We report two values here one the requires the cue match correctly and one that does not.
4. Negated Events: the metrics are computed only for negated events identification (apart from negation cue and scope).

Variant A								
	gold	system	tp	fp	fn	precision	recall	F1
Cues	264	250	232	14	32	94.31	87.88	90.98
Scope (cue match)	249	227	126	14	123	90.00	50.60	64.78
Scope (no cue match)	249	227	126	14	123	90.00	50.60	64.78
Scope (tokens - no cue match)	1805	1716	1456	260	349	84.85	80.66	82.70
Negated (no cue match)	173	183	70	70	64	50.00	52.24	51.10
Full negation:	264	250	75	14	189	84.27	28.41	42.49
Variant B								
	gold	system	tp	fp	fn	precision	recall	F1
Cues :	264	250	232	14	32	92.80	87.88	90.27
Scope (cue match):	249	227	126	14	123	55.51	50.60	52.94
Scope (no cue match):	249	227	126	14	123	55.51	50.60	52.94
Negated (no cue match):	173	183	70	70	64	38.25	52.24	44.16
Full negation :	264	250	75	14	189	30.00	28.41	29.18
# Sentences	1089							
# Negation sentences	235							
# Negation sentences with errors	171							
% Correct sentences	83.47							
% Correct negation sentences	27.23							

Table 3: Results of negation cue, negated event, and negation scope detection

5. Full negation: the metrics are computed for all the three tasks at once and requiring everything to match correctly.

For cue, scope and negated event to be correct, both the tokens and the words or parts of words have to be correctly identified. The final periods in abbreviations are disregarded. If gold has value "Mr." and system "Mr", system is counted as correct. Also, punctuation tokens are *not* taken into account for evaluation.

Two variants of the metrics are computed. In the first variant (A), precision is calculated as $tp / (tp + fp)$ and recall is calculated as $tp / (tp + fn)$ where tp is the count of true positive labels, fp is the count of false positive labels, and fn is the count of false negative labels. In variant B, the precision is calculated differently, using the formula $precision = tp / system$.

Table 3 shows the results of our system.

6 Error Analysis

The system used no external resources outside the training data. This means that the system recognizes only negation cues that appeared in the training set. This was the first source of error. For example, the word *unacquainted* that starts with the negation prefix *un* has never been seen in the training data. In-

tuitively, if no negation cue is detected, the system does not attempt to produce scope levels. This problem can be overcome by using a lexicon of negation words and those words that can be negated by adding a negation prefix to them.

We noticed in several occasions that scope detection accuracy can be improved if some simple rules can be imposed after doing the initial labeling using the CRF model (but we have not actually implemented any such rules in the system). For example, the system can require all the tokens that belong to the same chunk (noun group, verb group, etc.) all have the same label (e.g. the majority vote label). The same thing could be also applied on the segment rather than the chunk level where the boundaries of segments are determined by punctuation marks.

7 Conclusion

We presented a supervised system for identifying negation in English sentences. The system uses three CRF trained models. One model is trained for negation cue detection. Another model is trained for negated event identification. A third one is trained for negation scope identification. The models are trained using features extracted from a labeled dataset. Our experiments show that the system achieves promising results.

References

- Shashank Agarwal and Hong Yu. 2010. Biomedical negation scope detection with conditional random fields. *Journal of the American Medical Informatics Association*, 17(6):696–701.
- Wendy Webber Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, pages 301–310.
- Isaac G. Councill, Ryan McDonald, and Leonid Velikovich. 2010. What’s great and what’s not: learning to classify the scope of negation for improved sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, NeSp-NLP ’10, pages 51–59, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bas Heerschoop, Paul van Iterson, Alexander Hogenboom, Flavius Frasinca, and Uzay Kaymak. 2011. Analyzing sentiment in a large set of web data while accounting for negation. In *AWIC*, pages 195–205.
- Alexander Hogenboom, Paul van Iterson, Bas Heerschoop, Flavius Frasinca, and Uzay Kaymak. 2011. Determining negation scope and strength in sentiment analysis. In *SMC*, pages 2589–2594.
- Laurence R. Horn. 1989. *A natural history of negation / Laurence R. Horn*. University of Chicago Press, Chicago .:
- Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, April.
- Lifeng Jia, Clement Yu, and Weiyi Meng. 2009. The effect of negation on sentiment analysis and retrieval effectiveness. In *Proceedings of the 18th ACM conference on Information and knowledge management, CIKM ’09*, pages 1827–1830, New York, NY, USA. ACM.
- Jung-Jae Kim and Jong C. Park. 2006. Extracting contrastive information from negation patterns in biomedical literature. 5(1):44–60, March.
- 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.
- Roser Morante and Walter Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. *Proceedings of the Workshop on BioNLP BioNLP 09*, (June):28.
- Roser Morante, Anthony Liekens, and Walter Daelemans. 2008a. Learning the scope of negation in biomedical texts. *Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP 08*, (October):715–724.
- Roser Morante, Anthony Liekens, and Walter Daelemans. 2008b. Learning the scope of negation in biomedical texts. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 715–724, Honolulu, Hawaii, October. Association for Computational Linguistics.
- Roser Morante, Sarah Schrauwen, and Walter Daelemans. 2011. Annotation of negation cues and their scope. Technical report.
- Roser Morante. 2010. Descriptive analysis of negation cues in biomedical texts. *Language Resources And Evaluation*, pages 1–8.
- P. G. Mutalik, A. Deshpande, and P. M. Nadkarni. 2001. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. *Journal of the American Medical Informatics Association : JAMIA*, 8(6):598–609.
- Jonathon Read, Erik Velldal, Stephan Oepen, and Lilja vrelid. 2011. Resolving speculation and negation scope in biomedical articles with a syntactic constituent ranker. In *Proceedings of the Fourth International Symposium on Languages in Biology and Medicine*, Singapore.
- Gunnel Tottie. 1993. Negation in English Speech and Writing: A Study in Variation. *Language*, 69(3):590–593.
- Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. 2010. A survey on the role of negation in sentiment analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, NeSp-NLP ’10, pages 60–68, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT ’05*, pages 347–354, Stroudsburg, PA, USA. Association for Computational Linguistics.