An open-source tool for negation detection – a maximum-margin approach

Martine Enger and Erik Velldal and Lilja Øvrelid

Department of Informatics, University of Oslo
{marenger|erikve|liljao}@ifi.uio.no

Workshop on Computational Semantics Beyond Events and Roles
April 4 2017, Valencia, Spain

Negation cue and scope detection

Examples

- And yet it was **not** quite the last.
- Since we have been so **un**fortunate as to miss him and have [. . .]

Negation cue and scope detection

Examples

- And yet it was **not** quite the last.
- Since we have been so **un**fortunate as to miss him and have [...]

- Much published research, but hard to find available systems.
- Goal: Implement a lean and simple tool with minimal dependencies that
  - open and freely available,
  - easy to adapt (i.e. keep heuristics to a minimum),
  - draws on previous best practices, and
  - has competetive performance.
- Draws heavily on the design of the UiO_2 system (Lapponi et al., 2012) from the *SEM shared task 2012.

*SEM shared task

- We use the data and evaluation script from the 2012 *SEM shared task on negation detection (Morante & Blanco, 2012).
- Training, development and heldout testing based on the CoNLL-style Conan Doyle corpus (Morante & Daelemans, 2012).
- We only focus on cues and scopes (not events and focus).
- Use Stanford basic dependency representations rather than the provided constituent trees of the Charniak and Johnson (2005) parser.

System design, at a glance

- A maximum-margin learning approach for both cues and scopes.
- Implemented on top of PyStruct.
- Takes parsed (CoNLL-X) or raw text (assumes CoreNLP is installed).
- Cue detection: Binary SVM with lexical features
- Scope detection: SVM-based CRF with lexical and syntactic features.

Cue identification

- Closed-class assumption: only attempt to disambiguate cues seen during training (Velldal, 2011; Read et al., 2012).
- A lexicon of affixal cues is also automatically extracted
  - prefixes: {*dis, im, in, ir, un*}
  - infix: *less*
  - suffix: *less*
  - The classifier is presented with any words matching an affix pattern, e.g. *im*patient, *im*age and *im*aginary would match the prefix pattern.
- Features: Token PoS, form and lemma, as well as lemmas $\pm$ 1 position.
- Additional features for affixal candidates: the affix itself and character 5-grams from start/end of 'base'.
- Multi-word cues, e.g. '*by no means*' or '*neither...nor*': Post-processing.

Cue identification, results

	Development			Held-out		
	P	**R**	**F1**	**P**	**R**	**F1**
Baseline	90.68	84.39	87.42	87.10	92.05	89.51
UiO_2	93.75	95.38	**94.56**	89.17	93.56	91.31
System	91.67	95.38	93.49	90.15	93.56	**91.82**

- Majority class baseline:
  - Assign each word its most frequent label in the training data.
  - Outperforms 1/3 of the *SEM 2012 shared task systems.
- Slight drop in F1 when moving from the dev. to held-out set.
- Compared to UiO_2: recall is identical, but our system has more stable precision (1 percentage point higher on held-out).
- Would have ranked third in the *SEM 2012 shared task.

Scope resolution

- Approached as a sequence labeling task using a maximum-margin CRF.
- Features and labels inspired by the UiO_2 system from *SEM 2012.
- Configuration after tuning (along with the regularization parameter):

  Surface features: Form, lemma ($\pm$ 1), and PoS ($\pm$ 1), cue PoS, cue type, and left/right cue distance.

  Dependency features: Graph distance and path from cue.

  Label set: **B**eginning, **I**nside, **O**utside, and **C**ue.

Scope resolution results, for *gold cues*

- System comparison:
- Scope-level F1 for gold cues on the development and held-out set.

	Dev. F1	Test F1	Approach	Available
UiO_2	80.00	–	CRF	no
UiO_1	**82.52**	77.26	rules, SVM	no
Packard 2014	69.30	61.30	rules	scopes
Packard + UiO_1	82.50	**78.70**	rules×2, SVM	scopes, partly
Fancellu 2016	–	77.77	BiLSTM	scopes
System	77.38	77.26	CRF	scopes + cues

End-to-end results (scope-level metric)

	Development			Held-out		
	P	**R**	**F1**	**P**	**R**	**F1**
System (gold cues)	100.00	63.10	77.38	98.75	63.45	77.26
System	88.14	61.90	72.73	85.00	61.45	71.33

- Cue classification errors propagate to the scope classifier which will predict scopes for FP cues and do nothing for FN cues.
- Mostly affects precision.
- End-to-end system would have ranked 4th in the *SEM 2012 shared task w.r.t. the relevant subtasks.

Scope error analysis (for gold cues)

- Struggles with discontinuous scopes, as in:

  It was ***not****, I must confess, a very alluring prospect.*

- Other types of recurring errors: sentences with multiple negation cues with overlapping (gold) scopes.
- Many cases counted as FNs wrt the scope-level measure often just have a single token wrong, reflected in the higher token-level scores.

Future work

- Extend the scope resolution with post-processing heuristics for targeting discontinuous scopes.
- Train/test for other tasks, domains and annotation strategies.
- For example; speculation (and negation) detection based on BioScope.
- Pre-requisite: convert annotation format (XML $\rightarrow$ CoNLL-X).

Conclusion

- https://github.com/marenger/negtool
- A simple and open-source tool for detecting negation cues and their in-sentence scopes with competitive performance.
- Mostly relies on learned models, based on a maximum-margin approach.
- Pre-trained models for English are distributed along with the code, users can also train their own models.

Charniak, E., & Johnson, M. (2005). Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of the 43rd annual meeting of the association for computational linguistics* (pp. 173–180). Ann Arbor, MI, USA.

Lapponi, E., Velldal, E., Øvrelid, L., & Read, J. (2012). UiO_2: Sequence-Labeling Negation Using Dependency Features. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics (*SEM)* (pp. 319–327). Montreal, Canada.

Morante, R., & Blanco, E. (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)* (pp. 265–274). Montreal, Canada: Association for Computational Linguistic.

Morante, R., & Daelemans, W. (2012). ConanDoyle-neg: Annotation of negation in Conan Doyle stories. In *Proceedings of the First Joint*

*Conference on Lexical and Computational Semantics (*SEM)* (pp. 1563–1568). Montreal, Canada.

Packard, W., Bender, E. M., Read, J., Oepen, S., & Dridan, R. (2014). Simple Negation Scope Resolution through Deep Parsing: A Semantic Solution to a Semantic Problem. In *Proceedings of the 52nd annual meeting of the association for computational linguistics* (pp. 69–78). Baltimore, USA.

Read, J., Velldal, E., Øvrelid, L., & Oepen, S. (2012). UiO_1: Constituent-Based Discriminative Ranking for Negation Resolution. In *Proceedings of the first joint conference on lexical and computational semantics (*sem)* (pp. 310–318). Montreal, Canada.

Velldal, E. (2011). Predicting speculation: A simple disambiguation approach to hedge detection in biomedical literature. *Journal of Biomedical Semantics*, *2*(5).