Bridging the Gap Between Scope-based and Event-based Negation/Speculation Annotations:

A Bridge Not Too Far

Pontus Stenetorp*, Sampo Pyysalo,
Tomoko Ohta, Sophia Ananiadou and Jun'ichi Tsujii

http://pontus.stenetorp.se

<pontus@stenetorp.se>

Aizawa Laboratory
University of Tokyo

13th of July 2012

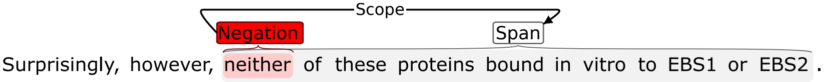
Outset

Outset

The Plan, Like So many Plans Before It

- Two separated research efforts
- A common goal
- Bringing them together in unification

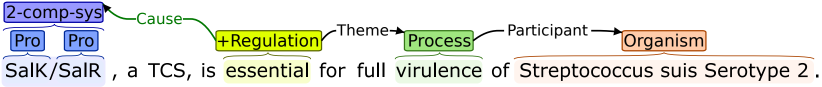
CoNLL-2010 Shared Task Negation/Speculation Corpus



Background

- Used in the CoNLL-2010 Shared Task (Farkas et al. 2010)
- Linguistically motivated cue-and-scope annotations

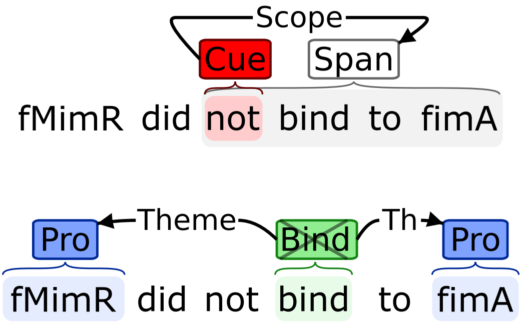
BioNLP Shared Tasks on Event Extraction (EE)



Background

- Introduced by Kim et al. (2009, 2011)
- Includes negation/speculation as a subtask
- Task-oriented binary flag annotations

Annotations Styles: Poles Apart?



Related Work

Kilicoglu and Bergler (2010)

- EE to extract cue-and-scope negation/speculation spans

Vincze et al. (2011)

- Manual analysis of negation/speculation annotation differences on the Genia EE and CoNLL-2010 corpus
- Found a multitude of problematic cases
- No proposed system to solve these issues

This Work

In a Single Question

- Can the efforts on the CoNLL-2010 Shared Task benefit those on event extraction?

This Work

In a Single Question

- Can the efforts on the CoNLL-2010 Shared Task benefit those on event extraction?

Approach

- Leverage existing systems for EE and negation/speculation
- Not duplicate previous efforts, join them
- Design inspired by manual analysis
- Propose a general solution

Data and Manual Analysis

Name	Neg. Events		Spec. Events		Neg. Spans	Spec. Spans
EPI	103	(5.6%)	70	(3.8%)	561	1,032
GE	759	(7.4%)	623	(6.0%)	1,308	1,968
ID	69	(3.3%)	26	(1.2%)	415	817

Table: Data sets from the BioNLP 2011 Shared Task (Kim et al. 2011)

Manual Analysis

- Vincze et al. (2011) analysis of GE
- Domain expert analysis of EPI and ID train sets

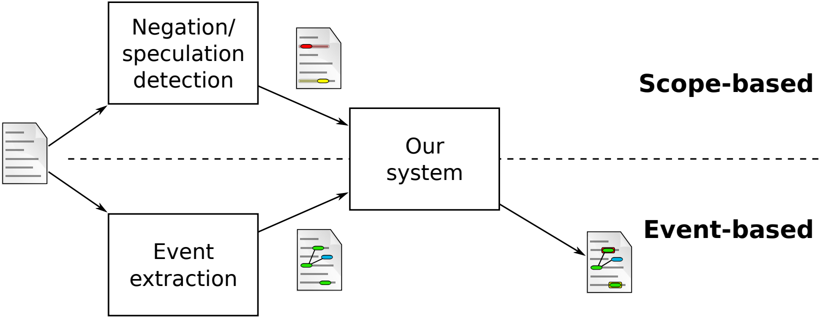
Manual Analysis Results

Analysis	Label	Ratio in EPI	Ratio in ID
	Covered	15.03	**56.52**
Coverage	Not-covered	**78.03**	41.30
	Error-in-gold	6.94	2.18

Manual Analysis Results

Analysis	Label	Ratio in EPI	Ratio in ID
Coverage	Covered	15.03	**56.52**
	Not-covered	**78.03**	41.30
	Error-in-gold	6.94	2.18
Non-covered	Morphological	**27.75**	11.96
	Hypothesis	25.43	**16.30**
	Ellipsis	2.89	0.00
	Argument-only	1.16	10.87

System Architecture



External Components

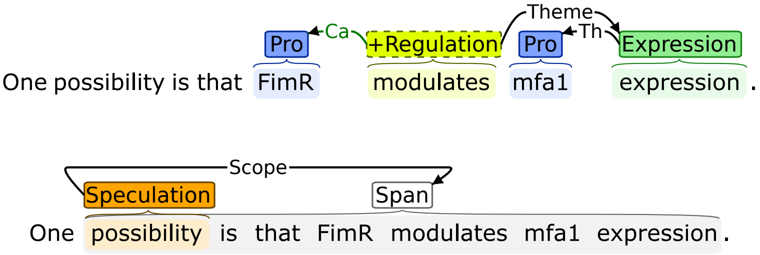
Negation/Speculation

- CLiPS-NESP (Morante et al. 2010)

Event Extraction

- UTurku (Björne et al. 2011)
- UConcordia (Kilicoglu and Bergler 2011)
- FAUST (Riedel et al. 2011)

Heuristic: Baseline



First Experiment

Set-up

- Negation/speculation from upstream system
- Gold event annotations
- Standard F-measure score

Heuristic: Negation Results

Negation	EPI	GE	ID
Heuristic-Baseline	30.40	53.38	36.97

Heuristic: Negation Results

Negation	EPI	GE	ID
Heuristic-Baseline	30.40	53.38	36.97
Heuristic-Root	30.00	61.18	40.74

Heuristic: Speculation Results

Speculation	EPI	GE	ID
Heuristic-Baseline	8.75	23.58	11.61

Heuristic: Speculation Results

Speculation	EPI	GE	ID
Heuristic-Baseline	8.75	23.58	11.61
Heuristic-Root	7.59	31.03	13.90

Machine Learning

Motivation

- Easier to incorporate complex features
- Re-trainable for different systems

Machine Learning

Motivation

- Easier to incorporate complex features
- Re-trainable for different systems

Machine Learning Procedures

- Tune penalty parameter using 10-fold cross-validation
- Merged training data
- Only final experiments on test set
- L2-regularised L2-loss SVM model

Machine Learning: Features

Group	Feature	Example Value(s)
Heuristic	Covered	ROOT/NON-ROOT
	Cue-Text	possibility
	Span-Tokens	One, possibility, ...

Machine Learning: Features

Group	Feature	Example Value(s)
Heuristic	Covered	ROOT/NON-ROOT
	Cue-Text	possibility
	Span-Tokens	One, possibility, ...
Trigger	Trigger-Text	non-phosphorylated
	Trigger-Prefixes	no, non, non-, ...

Machine Learning: Features

Group	Feature	Example Value(s)
Heuristic	Covered	ROOT/NON-ROOT
	Cue-Text	possibility
	Span-Tokens	One, possibility, ...
Trigger	Trigger-Text	non-phosphorylated
	Trigger-Prefixes	no, non, non-, ...
Contextual	Trigger-Preceding	is, that, ...
	Trigger-Proceeding	mfa1, expression, ...

Machine Learning: Negation Results

Negation	EPI	GE	ID
Heuristic-Baseline	30.40	53.38	36.97
Heuristic-Root	30.00	61.18	40.74
Trigger	28.18	31.82	33.82

Machine Learning: Negation Results

Negation	EPI	GE	ID
Heuristic-Baseline	30.40	53.38	36.97
Heuristic-Root	30.00	61.18	40.74
Trigger	28.18	31.82	33.82
Heuristic+Trigger	62.90	63.69	60.67

Machine Learning: Negation Results

Negation	EPI	GE	ID
Heuristic-Baseline	30.40	53.38	36.97
Heuristic-Root	30.00	61.18	40.74
Trigger	28.18	31.82	33.82
Heuristic+Trigger	62.90	63.69	60.67
Trigger+Context	52.00	66.15	56.52

Machine Learning: Negation Results

Negation	EPI	GE	ID
Heuristic-Baseline	30.40	53.38	36.97
Heuristic-Root	30.00	61.18	40.74
Trigger	28.18	31.82	33.82
Heuristic+Trigger	62.90	63.69	60.67
Trigger+Context	52.00	66.15	56.52
Heuristic+Trigger+Context	**64.96**	**70.84**	**63.74**

Machine Learning: Speculation Results

Speculation	EPI	GE	ID
Heuristic-Baseline	8.75	23.58	11.61
Heuristic-Root	7.59	31.03	13.90
Trigger	0.93	15.24	17.19

Machine Learning: Speculation Results

Speculation	EPI	GE	ID
Heuristic-Baseline	8.75	23.58	11.61
Heuristic-Root	7.59	31.03	13.90
Trigger	0.93	15.24	17.19
Heuristic+Trigger	5.88	29.00	28.57

Machine Learning: Speculation Results

Speculation	EPI	GE	ID
Heuristic-Baseline	8.75	23.58	11.61
Heuristic-Root	7.59	31.03	13.90
Trigger	0.93	15.24	17.19
Heuristic+Trigger	5.88	29.00	28.57
Trigger+Context	**52.43**	35.37	33.96

Machine Learning: Speculation Results

Speculation	EPI	GE	ID
Heuristic-Baseline	8.75	23.58	11.61
Heuristic-Root	7.59	31.03	13.90
Trigger	0.93	15.24	17.19
Heuristic+Trigger	5.88	29.00	28.57
Trigger+Context	**52.43**	35.37	33.96
Heuristic+Trigger+Context	49.50	**40.00**	**37.21**

Second Experiment: Enrichment

Set-up

- Negation/speculation from upstream system
- Event annotations from three upstream systems
- Global F-measure score
- EPI and ID data sets
- Heuristic+Trigger+Context model

Enrichment: Results

F-score	EPI	ID
UConcordia	27.88	44.21
UConcordia*	**28.05**	**45.19**

Enrichment: Results

F-score	EPI	ID
UConcordia	27.88	44.21
UConcordia*	**28.05**	**45.19**
UTurku	53.33	**42.57**
UTurku*	**54.29**	42.19

Enrichment: Results

F-score	EPI	ID
UConcordia	27.88	44.21
UConcordia*	**28.05**	**45.19**
UTurku	53.33	**42.57**
UTurku*	**54.29**	42.19
FAUST	35.03	55.59
FAUST*	**37.21**	**55.88**

Conclusions and Future Work

Conclusions

- Yes, we can bridge the gap!
- Introduced a practical way to do so
- Performed better than several existing EE systems
- Achieved state-of-the-art performance

Conclusions and Future Work

Conclusions

- Yes, we can bridge the gap!
- Introduced a practical way to do so
- Performed better than several existing EE systems
- Achieved state-of-the-art performance

Future Work

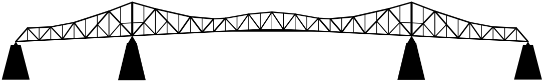
- Ways to eliviate hypothesis
- Other negation/speculation systems
- Use upstream confidence and/or integrate into EE system

Thank You for Your Attention

ご清聴ありがとうございました

Tack för er uppmärksamhet

Code: `http://ninjin.github.com/eepura/`
Slides: `http://pontus.stenetorp.se/`



Additional Negation Enrichment Results

Negation	EPI	GE	ID
UConcordia	26.51	25.88	22.92
UConcordia*	**31.17**	**27.42**	**29.68**
UTurku	18.60	**31.15**	**32.91**
UTurku*	**45.53**	27.33	32.10
FAUST*	39.18	28.25	**35.00**

Additional Speculation Enrichment Results

Speculation	EPI	GE	ID
UConcordia	**6.82**	**27.25**	3.23
UConcordia*	2.70	17.98	**3.51**
UTurku	37.65	**23.06**	**15.00**
UTurku*	**46.60**	15.60	4.76
FAUST*	41.76	14.93	12.50

Bibliography (1/3)

References

- Richárd Farkas, Veronika Vincze, György Móra, János Csirik, and György 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*.
- Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 Shared Task on Event Extraction. In *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*.
- Jin-Dong Kim, Sampo Pyysalo, Tomoko Ohta, Robert Bossy, Ngan Nguyen and Jun'ichi Tsujii. 2011. Overview of BioNLP Shared Task 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task*.

Bibliography (2/3)

References

- Halil Kilicoglu and Sabine Bergler. 2010. A High-Precision Approach to Detecting Hedges and their Scopes. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning.*
- Veronika Vincze, György Szarvas, György Móra, Tomoko Ohta and Richárd Farkas. 2011. Linguistic scope-based and biological event-based speculation and negation annotations in the BioScope and Genia Event corpora. *Journal of Biomedical Semantics.*
- Roser Morante, Vincent Van Asch and Walter Daelemans. 2010. Memory-based resolution of in-sentence scopes of hedge cues. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning.*

Bibliography (3/3)

References

- Jari Björne and Tapio Salakoski. 2011. Generalizing Biomedical Event Extraction. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task.*
- Halil Kilicoglu and Sabine Bergler. 2011. Adapting a General Semantic Interpretation Approach to Biological Event Extraction. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task.*
- Sebastian Riedel, David McClosky, Mihai Surdeanu, Andrew McCallum and Christopher Manning. 2011. Model Combination for Event Extraction in BioNLP 2011. In *Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task.*

For the Interested and Lawyers

Visualisations

- Annotation visualisations generated using the brat annotation/visualisation tool: `http://brat.nlplab.org/`

Legal

- Fractal cloud image from the Open Clipart Library
- Globe, Tank and bridge images from the Wikimedia Commons