



養天地正氣 法古今完人

# Negation and Speculation Target Identification

Bowei Zou, Guodong Zhou, Qiaoming Zhu

Soochow University



王舍人





# Outline

---

- Definitions and related work
- Target identification
- Corpus and annotation guidelines
- Method and experimentation





# What's a negation or speculation?



Philosophical perspective  
(Aristotle)



Typological perspective  
(Palmer 1986)



Cognitive perspective  
(Lawler 2010)

Logical perspective  
Psycholinguistic perspective  
.....



Linguistic perspective  
(Lakoff 1972)





# What's a negation or speculation?

## ➤ Two grammatical categories

- Negation
  - comprises various kinds of terms to reverse the truth value of a proposition
- Speculation
  - expresses the attitude of a speaker towards a statement in terms of degree of certainty, reliability, perspective, etc.

E.g.,

Our results show that **no** transcription of the RAG-1 gene could be detected.

The cardiovascular disease **may** recur even after cure.





# Why we study negation and speculation?



- Extract unreliable information which cannot be presented as factual information from clinical reports
- 11% of sentences in MEDLINE abstracts contain speculative fragments (Light et al. 2004)
- 18% of sentences in BioScope corpus contain speculation and 13% negation (Vincze et al. 2008)





# Existing Subtasks

- Existing research usually related to two subtasks
  - Cue detection
  - Scope resolution
- Cue (trigger/keyword)
  - A signal of a negative or speculative expression
- Scope
  - The grammatical coverage of cue in sentence

Our results show that [**no** transcription of the RAG-1 gene could be detected].

[The cardiovascular disease **may** recur] even after cure.





# Target Identification

- Which object is most exactly and prominently negated or speculated ?
- Target identification
  - Extract the object targeted by a negative or speculative expression.

Our results show that **no** transcription of the RAG-1 gene could be detected.

The cardiovascular disease **may** recur even after cure.





# Target Identification

## ➤ tow typical targets

- Entities directly modified by negative or speculative cues, ~40%
- Agents of verbal negative or speculative expressions, ~55% (patients in rare cases, <5%)

Our results show that **no** transcription of the RAG-1 gene could be detected.

The cardiovascular disease **may** recur even after cure.

The pharmaceutical plant is **difficult** to guarantee the drug to reach the safe quality requirements.







# Corpus

---

- Adding a target layer on BioScope (Vincze et al. 2008)
  - Biomedical literatures
  - Negative and speculative cues and scopes have been annotated
  - 1,668 of negation sentences and 2,678 of speculation
- Basic guideline
  - Target is the object which is most exactly and prominently negated or speculated in sentence





# Corpus

## ➤ Some specific guidelines

- Annotate in sentence

*It is **not** effective for all tuberculosis patients.*

- Maximal syntactic scope principle  
- including all adjuncts of target (prepositional phrases, determiners, adjectives)

*In contrast, blood lymphocytes **from patients with diseases** have **little** effect on children.*

- Extend to all members of a conjunction

*In common sense, symptoms include fever, cough or itches.*





# Corpus

## ➤ Annotation process

- ONLY refer to cues but NOT the corresponding scopes
- TWO annotators and ONE expert
- Inter-annotator agreement is 0.83 (kappa value)

	Negation	Speculation
#Sentence	1,668	2,678
In scope (%)	54.98%	63.71%
Out of scope (%)	45.02%	36.29%
Average length of sentences	29.73	31.16
Average length of targets	4.36	5.27
Before cue (%)	54.80%	43.05%
After cue (%)	45.02%	49.48%





# Method

---

## ➤ Feature-based Model

- All of noun phrases in sentence are taken as target candidates
- Basic features
  - candidate itself, cue itself/PoS
- Syntactic relationship(cue, candidate)
  - syntactic path, position, distance in paring tree, distance in sequence
- Neighbor(candidate)
  - left and right chunks/syntactic categories
- Semantic role
  - predicate, relationship between predicate and target
- Feature selection (Jiang et al 2006)





# Experiments

## ➤ Evaluation metrics

- Accuracy
  - Evaluating the performance of system
- Precision, Recall, and F-measure
  - Evaluating the performance of classifier

## ➤ Results

	P	R	F	Acc
Negation	76.27	63.53	69.32	70.13
Speculation	84.32	69.85	76.41	74.46

- Effective features
  - syntactic path between cue and candidate
  - candidate itself
  - right sibling tag of candidate's syntactic category
  - syntactic distance from candidate to keyword





## Conclusion

---

- Propose target identification, a new task on negation and speculation information extraction
- Add a layer of target information over BioScope
- Explore a set of effective features and presenting a benchmark system for future work





## Future work

---

- Establishing a complete framework for negation and speculation information extraction
  - Negation(Cue, Scope, Focus, Target, Holder, ...)
- Expanding the corpus to other domain, e.g., product review





Thanks







# Target Identification

## ➤ Vs. opinion target

- in Opinion Target Extraction (OTE) of sentiment analysis
- sentiment word  $\neq$  negative or speculative expression

## ➤ Vs. subject

- a constituent that conflates nominative case in sentence
- semantic perspective vs. syntactic perspective

The Prize of Best Employee is awarded **without** voting, unexpectedly.

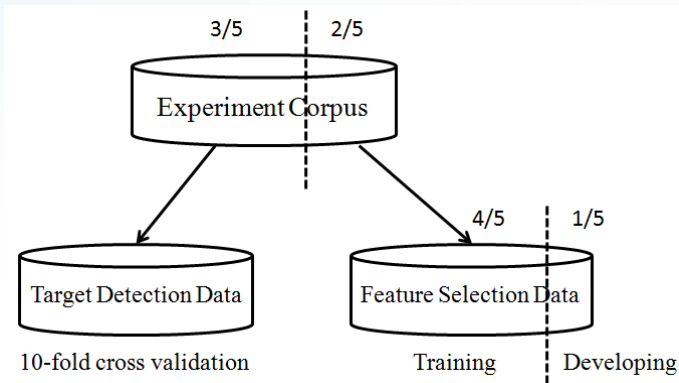
Company management has **not** yet decided on the Prize of Best Employee.





# Experiments

## ➤ Dataset



## ➤ Evaluation metrics

- Accuracy
  - Evaluating the performance of system
- Precision, Recall, and F-measure
  - Evaluating the performance of classifier





# Experiments

---

## ➤ Baselines

- Baseline\_First: the first noun phrase --> target
- Baseline\_Last: the last noun phrase --> target
- Baseline\_Longest: the longest noun phrase --> target
- Baseline\_Nearest: the nearest noun phrase from cue --> target  
(about nearest: E.g., in NP>S<VP<VBN, distance from NP to VBN is 3)

<i>Baseline</i>	Negation	Speculation
<i>First</i>	21.44	25.37
<i>Last</i>	27.41	24.19
<i>Longest</i>	32.13	35.59
<i>Nearest</i>	<b>39.93</b>	<b>43.77</b>





## Agent of an event

- A thematic relation that refers to the cause or initiator of an event (Kroeger, 2005, *Analyzing Grammar: An Introduction*. p.54)
- vs. Subject
  - Agent: determined through its relationship to the action expressed by the verb
  - Subject: determined syntactically, primarily through word order

The boy kicked the ball.  
Agent Subject

The ball was kicked by the boy.  
Subject Agent

