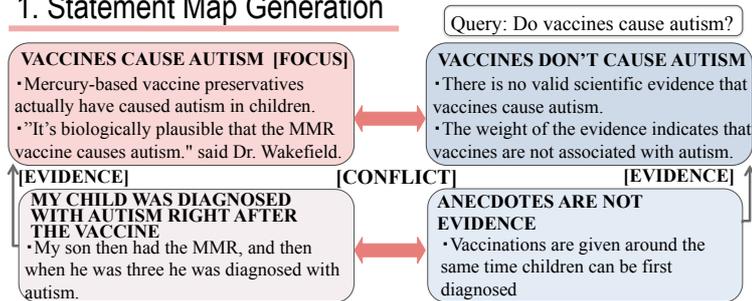


Annotating Semantic Relations Combining Facts and Opinions

Koji Murakami[†] Shouko Masuda^{†‡} Suguru Matsuyoshi[†]
Eric Nichols[†] Kentaro Inui[†] Yuji Matsumoto[†]
[†]Nara Institute of Science and Technology, Japan
[‡]Osaka Prefecture University, Japan

1. Statement Map Generation



Summary

- We defined semantic relations between facts and opinions.
- Relations are identified between sub-sentential fragments called **statements**.
- We annotated 1,600 Japanese sentence pairs with semantic relations.

Statement Map : Mapping arguments on the Web

We want to consider the CONTENTS of Web pages to evaluate the credibility of information on Web!!

- Automatically gather opinions and summarize them and organize into pros vs. cons
- Show users evidence supporting each position

Recognizing three major semantic relation classes

- [AGREEMENT]: to group similar opinions
- [CONFLICT]: to capture differences of opinions
- [EVIDENCE]: to show support for opinions

2. Facts and Opinions

A. RTE (Recognizing Textual Entailment) Challenge (Dagan et al, 2005)

-> Used to recognize logical/factual relations between sentence pair

B. CST (Cross Document Structure Theory) (Radev et al, 2001)

-> Used for objective expressions in newspaper articles

Under frameworks A/B, assigning any relation to the following pair is difficult

- (1-a) Mercury-based vaccines actually cause autism in children.
- (1-b) Vaccines can trigger autism *in a vulnerable subset of children*
- (2-a) There must not be a connection between vaccines and autism.
- (2-b) *I do believe* that there is a link between vaccinations and autism.

⇒ Define a wide spectrum of semantic relations to follow three relation classes, [AGREEMENT], [CONFLICT] and [EVIDENCE]

4. Definition of semantic relations

Relation Class	Relation Label	Description	Examples
AGREEMENT	Equivalence	Both A and B are TRUE at the same time	A: The overwhelming evidence is that vaccines are unrelated to autism. B: There is no link between the MMR vaccine and autism.
	Specific	Both A and B share the same information, and B has additional information	A: Mercury-based vaccine preservatives actually have caused autism in children B: Vaccines cause autism.
	Equivalent Opinion	Different source are in agreement or their opinions entail one another	A: We think vaccines cause autism. B: I am the mother of a 6 year old that regressed into autism because of his 18 month vaccinations.
	Equivalent Evaluative Polarity	A and B evaluate something from different perspective and their opinions have the same polarity	A: Vaccines are not effective. B: We think vaccines cause autism.
	Similar	A and B have similar sentence structure	A: MMR can cause autism. B: Mercury-based vaccines can cause autism.
CONFLICT	Contradiction	Both A and B can not be TRUE at the same time	A: Mercury-based vaccine preservatives actually have caused autism in children. B: Vaccines don't cause autism.
	Confinement	B confines the situations in which A applies	A: Vaccines can trigger autism in vulnerable subset of children. B: Mercury-based vaccine actually have caused autism in children.
	Conflicting Opinion	Different sources disagree or their opinions are contradictory	A: I don't think vaccines cause autism. B: I believe vaccines are the cause of my son's autism.
	Conflicting Evaluative Polarity	A and B evaluate something from different perspectives and their opinions have opposite polarities	A: We think vaccines cause autism. B: Vaccines are very important to protect our kids from dangerous disease like measles.
NO_RELATION	No Relation	There is no relation between A and B	A: In the UK, confidence in vaccines collapsed B: Parents should realize that a choice not to get a vaccine is not a risk-free choice

3. Constructing a Japanese Corpus

Real sentences in Web documents:

- complex structure => difficult to annotate semantic relations
- between parts of each sentence => able to annotate relations

→ Break sentences from the Web down into reasonable text fragments, which we call "**Statements**"

→ We label pairs of statements with a semantic relation or "No_Relation"

Preparing sentence pairs for annotation:

1. Extract sentences related to user's query from Web documents
2. Reduce the search space for identifying sentence pairs and prepare pairs, which look feasible to annotate

→ We filter sentences out with a method similar to (Dolan et al, 2005), and calculate the lexical similarity between two sentences based on BOW

5. Statistics

- 5 people annotated relations in 22 document sets
- Provide 2,303 real sentence pairs to annotators
- 928 pairs were identified as "valid"
- 1,612 statement pairs were annotated (AGREEMENT:890(55%), CONFLICT:222(14%), NO_RELATION:500(31%))
- 81.6% of agreements between annotators for 207 randomly selected statement pairs (corresponds to a kappa level of 0.49)

		Annotator-A			
		AGR.	CON.	NONE	TOTAL
Annotator-B	AGR.	146	7	9	162
	CON.	0	13	1	14
	NONE	17	4	10	31
	TOTAL	163	24	20	207

6. Conclusion and Future Work

- Designed an annotation scheme with the necessary 9 semantic relations
- Annotated 1,612 statement pairs with a semantic relation or "No_Relation"
- Achieved 81.6% of inter-annotator agreement
- Constructing a corpus for the remaining semantic relation, "EVIDENCE"
- Annotating relations for more than 6,000 sentence pairs in this autumn.
- Releasing the corpus (<http://stmap.naist.jp/corpus/ja>)

Acknowledgement : This work is supported by the National Institute of Information and Communications Technology Japan.