

## Overview

Engaging in a live debate requires rebutting opponent arguments

**What are those arguments?**

Corpus-wide claim mining is a fundamental capability of Project Debater

**Mine claims from the opponent stance**

Will mined claims be mentioned by opponents?

**How can they be identified, when listening live?**



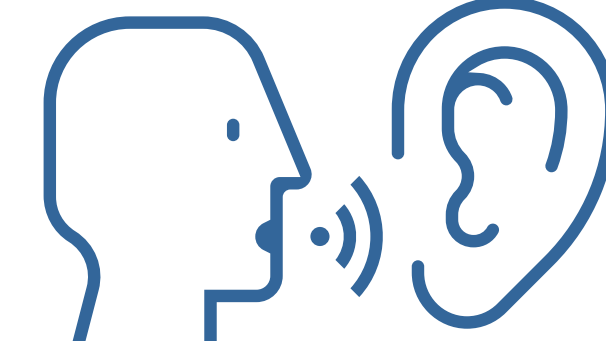
Opponent  
Speech  
Audio



Massive corpus  
~10B sentences



A 2017 study showed  
that **water fluoridation**  
has no side-effects



Listening  
comprehension



Mentioned  
claims

## Annotation example

**Speech:**

*"It's essential that something is done to ensure that people don't have dental problems later in life. Water fluoridation is so cheap it's almost free. There are no proven side effects, the FDA and comparable groups in Europe have done lots and lots of tests and found that water fluoridation is actually a net health good, that there's no real risk to it "*

**Mined claim:**

*Water fluoridation has no side-effects*

*Water fluoridation is safe and effective*

*Most people wouldn't get regular fluoride treatment at the dentist*

**Label:**

*Explicit*

*Implicit*

*Not mentioned*

## Debater's corpus-wide claim mining pipeline

Sentence-level query



Sentence ranker



Boundary detection



Stance detection

*Levy et al. '14,'17,'18*

*Bar-Haim et al. '17*

**Topics:**

**Speeches:**

**Topic coverage:**

**Annotated claims:**

**Implicit:**

**Explicit:**

**New dataset**

200

400

93.5%

4,882

34.7%

5.6%

**Mirkin '18**

50

200

39%

878

40.3%

## Annotation quality

Avg. inter-annotator agreement (Kappa): 0.44

Avg. annotator test error rate: 7.8%

Platform: Figure-Eight

## Get the dataset

- Audio
- ASR
- Transcripts
- Labeled claims



## Listening comprehension benchmarks

**Goal: Identify mined claims mentioned in an opponent's speech**

-Use IBM Watson speech-to-text service

-Filter out sentences with low w2v similarity

-Identify the K sentences that are most semantically similar to the claim

-Experiment with 3 baselines:

Unsupervised (**HM**) | Logistic Regression (**LR**) | Siamese network (**NN**)

**Benchmark**      **AUC**

HM      0.66

LR      0.56

NN      0.52

## Future work

**Goal: Rebut an identified claim persuasively**

-Find a counter-argument in

the text corpus

-Generation

-Prewritten general purpose

response