CrowdTruth Metrics for Capturing Ambiguity

Interlinking Workers, Annotations and Input Data

SAD workshop @ HCOMP
July 5 2018

VU UNIVERSITY AMSTERDAM

Anca Dumitrache, Oana Inel, Lora Aroyo, Chris Welty

Traditional Human Annotation

- Humans provide annotations establishing the **ground truth** = the correct output for each example (the **gold standard**)

- **Machines learn** the ground truth

- **Ground Truth** Quality: typically measured by inter-annotator agreement (e.g. majority vote); founded on **the ideal for single, universally constant truth**

- which means - **ambiguity** of textual interpretation is often lost

Anca Dumitrache
@anca_dmtrch

CrowdTruth Methodology

- Annotator disagreement is signal, not noise.

- It can indicate of the variation in human semantic interpretation

- Can be used to capture ambiguity, vagueness, similarity, over-generality, as well as quality

What causes disagreement to happen?

Anca Dumitrache
@anca_dmtrch

Disagreement because of Low Quality Workers

Do the sentences express a **treat** relation?

`ANTIBIOTICS` are the first line treatment for vectors of `TYPHUS` .

`QUININE` is not a reliable cure for `MALARIA` .

Worker	Sentence (1)	Sentence (2)
Worker 1	yes	no
Worker 2	yes	no
Worker 3	yes	
Worker 4		no
Worker 5	no	yes

Anca Dumitrache
@anca_dmtrch

Disagreement because of Sentence Clarity

Do the sentences express a **treat** relation between CHLOROQUINE and MALARIA ?

Rheumatoid arthritis and MALARIA have been treated with CHLOROQUINE for decades.

For prevention of malaria, use only in individuals traveling to malarious areas where CHLOROQUINE resistant P. falciparum MALARIA has not been reported.

Among 56 subjects reporting to a clinic with symptoms of MALARIA 53 (95%) had ordinarily effective levels of CHLOROQUINE in blood.

Anca Dumitrache
@anca_dmtrch

Disagreement because of Sentence Clarity

Do the sentences express a **treat** relation between CHLOROQUINE and MALARIA ?

Rheumatoid arthritis and MALARIA have been treated with CHLOROQUINE for decades.
→ agreement 95%

For prevention of malaria, use only in individuals traveling to malarious areas where CHLOROQUINE resistant P. falciparum MALARIA has not been reported.
→ agreement 75%

Among 56 subjects reporting to a clinic with symptoms of MALARIA 53 (95%) had ordinarily effective levels of CHLOROQUINE in blood.
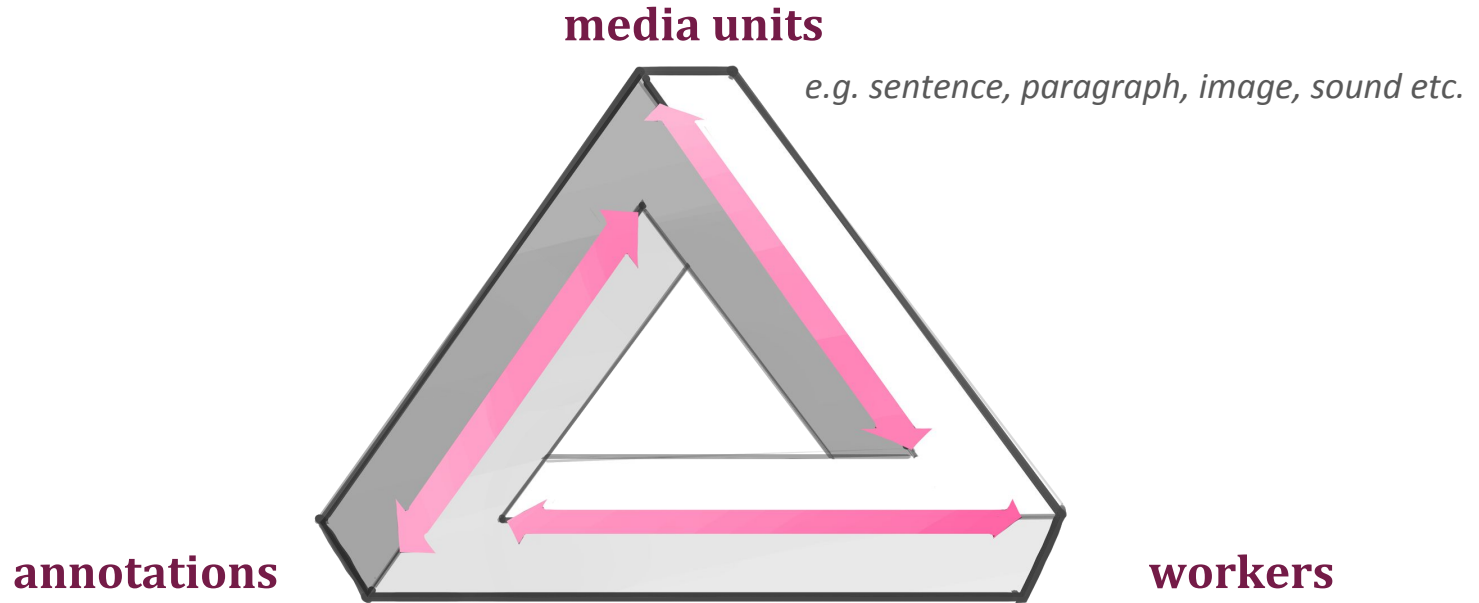→ agreement 50%

Anca Dumitrache
@anca_dmtrch

Disagreement because of an Ambiguous Annotation Task

METHYLERGOMETRINE is a blood vessel constrictor most commonly used to control excessive BLEEDING .

What is the relation expressed?
Treat or **Prevent**?

Anca Dumitrache
@anca_dmtrch

Triangle of disagreement as model for crowdsourcing systems

media units

e.g. sentence, paragraph, image, sound etc.

annotations

workers

Ambiguity at any corner disseminates in the other corners

Anca Dumitrache
@anca_dmtrch

CrowdTruth quality metrics

- **Media unit quality:** overall worker agreement over one media unit
- **Worker quality:** overall agreement between one worker and the other workers
- **Annotation quality:** overall worker agreement over one annotation

Calculated using *average cosine similarity* over media unit vectors.

Quality metrics are ***inter-dependent***.

e.g. ambiguous *media units* count less when determining *worker* quality, because disagreement is expected

Anca Dumitrache
@anca_dmtrch

CrowdTruth.org

github.com/CrowdTruth/CrowdTruth-core

pypi.org/project/CrowdTruth

data.CrowdTruth.org