The Search for Agreement on Logical Fallacy Annotation of an Infodemic

LREC 2022
Claire Bonial, Austin Blodgett, Taylor Hudson, Stephanie M. Lukin, Jeffrey Micher, Douglas Summers-Stay, Peter Sutor, & Clare R. Voss

Tools for Combating Misinformation

What are the tools in our toolbox?

Tools for Combating Misinformation

What are the tools in our toolbox?

NLP Applications where Misinformation is a relevant problem:
- Document Retrieval
- Question Answering
- Summarization

Tools for Combating Misinformation

What are the tools in our toolbox?

With External Knowledge
- Fact-Checking
  Automatically (or by hand) check claims against external information
- White and Black Lists
  Categorize websites and sources by their reputation for credible/non-credible information

Tools for Combating Misinformation

What are the tools in our toolbox?

What about fast changing topics (e.g., during a pandemic)?

Tools for Combating Misinformation

What are the tools in our toolbox?

What about fast changing topics (e.g., during a pandemic)?

Without External Knowledge
- Rumor Detection
  Identify messages that spread quickly
- Propaganda Detection
  Identify strategies for misleading readers
- Fallacy Detection
  Identify flawed arguments

Envisioned System

User:

Do I need to sanitize my mask?

System:

Here's an article claiming the importance of proper mask care. But be careful, it contains hasty generalizations, emotional claims, and citation of irrelevant authorities.

Appeal to Emotion   Irrelevant Authority

Wash your mask. **Wake up!!!**   **A caller to a radio talk show**

recently shared that his wife was hospitalized and told she

had COVID and only a couple of days left to live. A doctor

friend suggested she be tested for legionnaires disease because

she wore the same mask every day all day long. Turns out it

WAS legionnaires disease from the moisture and bacteria in her

mask. She was given antibiotics and within two days was better.

WHAT IF these ""spikes"" in COVID are really something else

due to improper mask wearing??

Hasty Generalization

In the long term, we want systems like this one, where users can ask questions and a system retrieves relevant documents but with warnings about possible fallacies.

Envisioned System

User:
Do I need to sanitize my mask?

System:

Here's an article claiming the importance of proper mask care. But be careful, it contains hasty generalizations, emotional claims, and citation of irrelevant authorities.

Appeal to Emotion | **Irrelevant Authority**

Wash your mask. **Wake up!!!** **A caller to a radio talk show recently shared that his wife was hospitalized and told she had COVID and only a couple of days left to live.** A doctor friend suggested she be tested for legionnaires disease because she wore the same mask every day all day long. Turns out it WAS legionnaires disease from the moisture and bacteria in her mask. She was given antibiotics and within two days was better. **WHAT IF these ""spikes"" in COVID are really something else due to improper mask wearing??**

Hasty Generalization

In the long term, we want systems like this one, where users can ask questions and a system retrieves relevant documents but with warnings about possible fallacies.
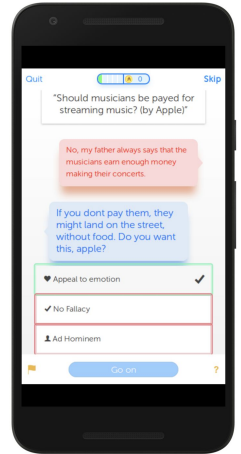
Where to get the data?

Annotation Task

Argatario: Crowdsource annotations of fallacies through a gamified task.

- User Created Data
- "Gold" labels were agreed on by a multiple users

Annotation Task

Argatario: Crowdsource annotations of fallacies through a gamified task.

- User Created Data
- "Gold" labels were agreed on by a multiple users

Our Work: Linguist annotated fallacies

- Data gathered from multiple sources on topics related to COVID-19
- Data double-annotated and then adjudicated by linguist annotators

Corpus Composition

Topic	Genre (# of docs)
COVID Vax Safety	Online Medical Forum (2)
	Tabloid (1)
	Science Magazine (1)
	Social Media (1)
Herd Immunity	General News (3)
	Talk Radio (1)
Long Haulers	Online Medical Forum (4)
Mask Safety	General News (2)
	Social Media (3)
General Vax Safety, Efficacy	General News (1)
	Health Care Site (3)
SARS-CoV-2 Origin	General News (2)
	Scientific Article (2)

Annotation Task

- Ad Hominem:
  - *It's just too convenient for vax-pimping scientists to claim that their precious vaccines don't work because not enough people are getting them* (Topic: General vaccine safety and efficacy)
- Appeal to Emotion:
  - *It is time for families to wake up to uncloaking of the new world order in its glory.* (Topic: COVID-19 vaccine safety and efficacy)
- Red Herring:
  - *Being a real scientist would be easy if it weren't for this `needing evidence' stuff, just like being a professional golfer would be simple if it weren't for this `having to put the ball in the hole thing.'* (Topic: SARS-CoV-2 virus origin)
- Hasty Generalization:
  - *They're not reporting the number of deaths per million. In other words, they're not reporting the survivability rate ... The answer here is don't mandate closures.* (Topic: Herd immunity)
- Irrelevant Authority:
  - *'Inside Edition' also lauded Biden, Mitt Romney, and Tom Cruise for double masking recently.* (Topic: Mask safety and efficacy)
- Or *None of the Above*

Results

We evaluated Inter-annotator agreement for two decision levels:

- L1: Whether or not there was a fallacy (*None of the Above* vs. *Fallacy Label*)

- L2: Which fallacy was present (Only for sentences where both annotators annotated a fallacy)

Results

Data	Level 1 IAA	Level 2 IAA
Full Corpus	.47	.51
First-round	.46	.54
Second-round/Test	.51	.31
General News	.23	-
Health Care Sites	.48	-
Online Med Forum	.26	-
Talk Radio	.48	-
Science Magazine	.31	-
Scientific Article	.33	-
Tabloid	.13	-

Additional Experiment: PET fallacy prediction

Pattern Exploiting Training (PET) is a procedure for tuning a pre-trained LM for classification tasks that it wasn't designed to solve, by leveraging cloze-style questions.

We experiment with using PET to predict fallacies based on our gold data.

Additional Experiment: PET fallacy prediction

*Consider the passage [**insert passage**]. Which of the following fallacies is this an example of? a) Hasty Generalization, b) Appeal to Emotion, c) Red Herring, d) Ad Hominem, e) Irrelevant Authority, f) None of the Above. The correct answer is _____.*

Hasty Generalization
Appeal to Emotion
Red Herring
Ad Hominem
Irrelevant Authority
None of the Above

Additional Experiment: PET fallacy prediction

Data / Agreement Level	A1 vs. A2	A1 vs. Gold	A2 vs. Gold	PET vs. Gold
Gold Corpus / Level 1	.46	.71	.82	-
Gold Corpus / Level 2	.54	.45	.56	-
Test / Level 1	.51	.77	.73	.22
Test / Level 2	.31	.72	.38	-.07

Challenges & Lessons

We identify 3 major sources of ambiguity in when doing fallacy annotation:

Challenges & Lessons

We identify 3 major sources of ambiguity in when doing fallacy annotation:

1. Linguistic Anchor

Challenges & Lessons

We identify 3 major sources of ambiguity in when doing fallacy annotation:

1. Linguistic Anchor

2. Overlapping Definitions

Challenges & Lessons

We identify 3 major sources of ambiguity in when doing fallacy annotation:

1. Linguistic Anchor

2. Overlapping Definitions

3. External Knowledge

Challenges & Lessons: Linguistic Anchor

The anchor for a fallacy might be a sentence, clause, phrase, or multiple sentences. When a fallacy existed in a context of multiple sentences, annotators had difficulty identifying the right anchor.

"Israel Today newspaper announced that Saudi Arabia is a close ally. This followed an announcement that Saudi Arabia, the UAE, Egypt and Bahrain will boycott Qatar, citing its support for terrorism." - Red herring, but which sentence do we label?

One of these is a premise and one is the conclusion of a fallacy.

Challenges & Lessons: Overlapping Definitions

Overlapping Definitions

Appeal to Emotion vs. Hasty Generalization

"I suspect these draconian organized crackdowns on health freedom will become a permanent reality. Government will use the Corona scare as a pretext to fast-track vaccine mandates into law everywhere."

Appeal to Emotion vs. Red Herring

"It is never wise to defer you personal decisions to an external authority, but especially now. This dilemma has already redefined the landscape, giving rise to new authoritarianism."

Challenges & Lessons: Overlapping Definitions

	HAST	EMOT	RED	AD	IRR
HAST	16	14	6	0	0
EMOT		40	13	3	1
RED			19	1	2
AD				3	0
IRR					2

Challenges & Lessons: External Knowledge

Some fallacy judgements require external knowledge that a system will not have at run time. Annotators will also have different domain specific knowledge and annotate differently as a consequence.

Irrelevant Authority

- Is John/Jane Doe a relevant authority on X. We don't know without more context.

Red Herring

- Is "cytokine" relevant to discussions of COVID-19? We need Domain-specific knowledge.

Conclusions & Future Work

Research on Fallacy Detection needs to be able to answer the following questions:

- What is the right linguistic anchors for fallacy annotation?
- What granularity of fallacy labels should be used?
- How do we handle (or weed out) external knowledge?
- Can we make litmus tests to evaluate more objectively if a given fallacy is present?

Thank You