Natural Language Processing for Sustainable Diets

Research Seminar at CApE
2 November 2022

Daniel Hershcovich
Department of Computer Science (DIKU)
What is Natural Language Processing?
Language models

Instruction finetuning
Please answer the following question. What is the boiling point of Nitrogen?

-320.4F

Chain-of-thought finetuning
Answer the following question by reasoning step-by-step.
The cafeteria had 23 apples. If they used 20 for lunch and bought 6 more, how many apples do they have?

The cafeteria had 23 apples originally. They used 20 to make lunch. So they had 23 - 20 = 3. They bought 6 more apples, so they have 3 + 6 = 9.

Multi-task instruction finetuning (1.8K tasks)
Inference: generalization to unseen tasks

Q: Can Geoffrey Hinton have a conversation with George Washington? Give the rationale before answering.

Geoffrey Hinton is a British-Canadian computer scientist born in 1947. George Washington died in 1799. Thus, they could not have had a conversation together. So the answer is “no”.

Scaling Instruction-Finetuned Language Models (Chung et al., 2022)
**Language models**

**Input text**

Write a sad story about a carrot named Jason. The story should start with the carrot being a professional athlete of some kind, and end with the carrot having his heart broken.

**Flan-PaLM output**

Jason Carrot, the famous professional athlete, had it all. He had the good looks, the fame, the fortune, and the women. Well, he thought he had it all. Jason was married to a beautiful girl, Betty Beets. Or so he thought. One day, Jason came home from a press conference to find Betty in bed with another man. Jason was crushed, and he fell into a deep depression. He stopped playing professionally, he lost his fortune, and he gained weight. Jason is now a nobody, and he has no girlfriend.

**Input text**

Write me a funny poem about a cat driving a car.

**Flan-PaLM output**

I’ve seen a cat driving a car.
I saw it on my way to the bar.
I stopped to look and say hi.
I asked if he could give me a ride.
He looked at me and said no.
Then he drove off very slow.
I wonder where he was going.
Maybe he was just going bowling.

**Scaling Instruction-Finetuned Language Models** (Chung et al., 2022)
Misinformation detection

Increasing dietary magnesium intake is associated with a reduced risk of stroke, heart failure, diabetes, and all-cause mortality.

The study findings suggest that increased consumption of magnesium-rich foods may have health benefits.

#Magnesium saves lives  https://t.co/K0M6QdjWcc
https://t.co/FUtWZ8jADM

Modeling Information Change in Science Communication with Semantically Matched Paraphrases (Wright et al., EMNLP 2022)
Arguments mining and generation

Argument mining
Given a topic, extract relevant arguments from large text corpus

**Topic:** *We should abandon Valentine's day*

The Canadian polling firm Insights West *surveyed* a representative sample of Canadians who are in a relationship and found that *62 percent* agreed that *Valentine's Day* is a waste of time and *money*.

A recent *survey* by Virgin Mobile USA found that *59 percent* of people said that if they were going to break up with someone, they would do so just before *Valentine's Day* to save *money*.

*Advances in Debating Technologies: Building AI That Can Debate Humans* (Bar-Haim et al., ACL 2021)
An autonomous debating system
Slonim et al. (Nature 2021)
Discovering semantic and cultural change

*Mining the Meaning: A semantic parsing approach to the study of literary change*

Data+ project with Jens Bjerring-Hansen and Bolette S. Pedersen, Department of Nordic Studies and Linguistics

Measuring semantic change in 19\textsuperscript{th}-century Scandinavian novels

Postdoc starting November 15\textsuperscript{th}
What does this have to do with sustainability?
Climate impact of NLP

Models require more and more resource and energy

Efficient Methods for Natural Language Processing: A Survey (Treviso et al., 2022)
Climate awareness in NLP

- Energy efficient NLP algorithms and methods are becoming widespread

- Efficient is not enough – net impact on the environment should be positive

- Example applications include verifying compliance, modeling consumer behavior

Towards Climate Awareness in NLP Research
(Hershcovich et al., EMNLP 2022)
Historical and cultural perspective

Average amount of fruits and vegetables available per person per year (kg)

Source: WHO
Denmark's Official Dietary Guidelines

- *Eat plant-rich, varied and not too much*
- *Eat more vegetables and fruit*
- *Eat less meat – choose legumes and fish*
- *Eat wholegrain foods*
- *Choose vegetable oils and low-fat dairy products*
- *Eat less sweet, salty and fatty food*
- *Thirsty? Drink water*
Enriching Arguments with Diverse Perspectives

• Denmark's Official Dietary Guidelines argues "for health and climate"
• What about cultural, economic, ethical perspectives?
Argumentation

Can AI convince consumers of diverse backgrounds to change their habitual behavior?

What characterizes successful arguments for behavior change in public discourse?
Argument mining on social media

1. Corpus of 32,840 English tweets
2. 600 arguments annotated for 5 topics
3. Trained and evaluated argument mining models

Topics:
- We should reduce the consumption of meat
- Plant based food should be encouraged
- Meat alternatives should be encouraged
- Vegan and vegetarian diets should be encouraged
- We should pursue policies that promote sustainable foods

A Dataset of Sustainable Diet Arguments on Twitter (Hansen & Hershcovitch, NLP4PI 2022)
Annotation methodology

Labels:

Argumentative - text contains argumentative structures such as claims or evidence while having a clear stance towards a topic

Claim – a standpoint towards the discussed topic implicitly or explicitly

Evidence - statement used to support or attack a topic or claim

Pro/con - the stance of a claim or the evidence towards the discussed topic
Types of evidence

Anecdotal – A description of an episode(s), centered on individual(s) or clearly located in place and/or in time

Expert – Testimony by a person, group, committee, an organization with some known expertise/authority on the topic

Study – Results of a quantitative analysis of data, given as numbers, or as conclusions

Normative – Added description for a belief about the world

Fact – A known piece of information about the world without a clear source for the information
Annotation methodology

Crowdsourced Annotations

- 597 tweets annotated for 5 topics
- Using Amazon Mechanical Turk
- Completed in four rounds
Topic: Plant based food should be encouraged

Tweet: <MENTION> Many fruit & veg (which contain natural acid) don’t trigger flare ups- The list is long and varied (obs this may not apply to you) but after a little digging I found some doctors do recommend a plant based diet to ease the inflammation. Going meatless is even recommended by ICA

Label: Claim, Evidence and Pro

Topic: Meat alternatives should be encouraged

Tweet: <MENTION> People need to stop trying to recreate meat without the…ya know meat

Label: Claim and Con
Examples

Topic: We should reduce the consumption of meat

Tweet: Green taxes go into subsidizing development and production of green energy solutions. If we were on 100% renewables, our electricity prices would not have needed to go up. We need to move into self-sufficient green energy as soon as possible

Label: Argumentative but unrelated

Topic: We should reduce the consumption of meat

Tweet: <MENTION> <MENTION> Lol - and the wash post is the PR firm and Whole Foods is the official food supplier

Label: Not argumentative
### Diverse perspectives

<table>
<thead>
<tr>
<th>Tweet</th>
<th>Aspects</th>
<th>Stance</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Many fruit &amp; veg (which contain natural acid) don’t trigger flare ups- The list is long and varied (obs this may not apply to you) but after a little digging I found some doctors do recommend a plant based diet to ease the inflammation. Going meatless is even recommended by ICA</td>
<td>Health</td>
<td>Positive</td>
</tr>
<tr>
<td>(2) People have been making meat alternatives for centuries for various reasons. If people want to eat a meat analogue that causes much less death and environmental destruction, what’s your problem?</td>
<td>Culture, ethics</td>
<td>Positive</td>
</tr>
<tr>
<td>(3) Fair enough tbh, I can’t become vegan or vegetarian due to personal reasons but the way animals are treated in the food industry is absolutely disgusting</td>
<td>Lifestyle, ethics</td>
<td>Positive</td>
</tr>
<tr>
<td>(4) Switch to plant based diets to reduce food costs and benefit climate</td>
<td>Economy, climate</td>
<td>Positive</td>
</tr>
<tr>
<td>(5) More people employed in plant based food production</td>
<td>Economy</td>
<td>Positive</td>
</tr>
<tr>
<td>(6) Yes but to be fair: we can expect a massive increase in meat and dairy consumption in emerging countries that will severely limit the impact of whatever we do</td>
<td>Economy</td>
<td>Negative</td>
</tr>
</tbody>
</table>
Potential applications

- Surveying public attitudes and exposing misinformation
- Generation of convincing arguments for communication campaigns
- Understanding public knowledge of sustainable diets
Facilitate Systematic Understanding

Mapping out consumer opinions

• Social media
• Local groups
Thank you

danielhers.github.io
dh@di.ku.dk
Addressing Social Bias

Different arguments are convincing for different social groups
Cultural awareness in NLP

Objectives and Values

Linguistic Form and Style

Common Ground

Aboutness

Challenges and Strategies in Cross-Cultural NLP
(Hershcovich et al., ACL 2022)
How we express ourselves in language

Morphosyntax
Word choice
Style
Levels of granularity

Linguistic and cultural variation within groups

<table>
<thead>
<tr>
<th>Idiolect</th>
<th>Sociolect, dialect</th>
<th>Standardised language</th>
<th>Language, language family</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual,</td>
<td>Social group or</td>
<td>Country, national</td>
<td>International cultures</td>
</tr>
<tr>
<td>personality</td>
<td>region, sub-culture</td>
<td>culture</td>
<td></td>
</tr>
</tbody>
</table>
Common ground

Shared knowledge based on which people reason and communicate

- Conceptualisation
- Commonsense
Commonsense

Some knowledge is "universal", other culture-specific

GeoMLAMA: Geo-Diverse Commonsense Probing on Multilingual Pre-Trained Language Models
(Yin et al., EMNLP 2022)
Aboutness ❤️

What content do people *care about*?
Values

Objectives and goals people strive for

What is considered desired or desirable

Cultures and Organizations: Software of the Mind (Hofstede, 1991)