Structured Argument Mining in Persuasive Online Articles on Sustainable diets

By Lukas Mikelionis, MSc in Computer Science at UCPH.
Supervised by Asst. Prof. Daniel Hershcovich
In collaboration with CONCITO

In order to tackle pressing issues such as global warming and the food crisis, it is necessary to explore alternative dietary choices that are sustainable and eco-friendly. This entails addressing the overconsumption of animal products. However, influencing society's eating habits to make a significant impact is a difficult undertaking due to deeply ingrained food traditions, misconceptions regarding plant-based diets and the lack of effective communication from the authorities.

Luckily, in today's digital age, the Internet is full of content which could be utilized exactly for the purpose of encouraging behavioral change in society. Of course, the vast amounts of data and its rapid growth have their challenges as well, as it is simply impossible for people to keep track of everything. This is where Argument Mining (AM), a sub-field of Natural Language Processing (NLP), comes in to automatically extract relevant information from texts, more specifically, to find arguments in various types of media pieces like articles and tweets. For instance, the sentence "*Combining cultivated and plant-based meat is unlikely to be the key to consumers craving slaughter-free protein.*" [1] is a claim that could be potentially argued in favor or against by different individuals.

The goal of the project was to collect a dataset of persuasive online articles that could then be employed in training Artificial Intelligence (AI). The expectation is that the AI would "learn" to automatically extract argumentative information, such as claims and evidence, from the text. This novel combination of the task, structured argument mining, and media type, persuasive online articles, had not been explored in the existing literature. Structured argument mining here means not only identifying argumentative excerpts in the text but also connections between them. For instance, the sentences "*Aside from a lack of infrastructure in place to support global cultivated production, **consumer perceptions could also stunt sector progress**.*" and "*Last year, an Australian study reported that **72 percent of diners are not willing to consider cultured meat yet**.*" [1] are clearly linked (argumentative excerpts are in bold): the first makes an assertion while the second provides evidence. Furthermore, opinion articles, in particular, serve as a natural source of arguments for practical applications because their primary objective is to convince readers or, at the very least, prompt them to consider alternative viewpoints.

The collected corpus consists of 20 opinion articles on the topic of sustainable diet, published on three well-known news websites: Altinget, Plant Based News and the Guardian. For the AI to be able to perform a task, it first needs to see some reference solutions for it to "learn" from. To accommodate this, arguments in the 20 articles were

annotated by three individuals who participated in the study. In total, 1845 argumentative components (including claims and premises) were identified by the three individuals. About 65% of the text in articles was considered as containing arguments, showing that even opinion articles have a considerable amount of content that is dedicated for other purposes besides persuading the readers(i.e., providing factual information). The participants of the study were compensated using the funds provided by the Green Solutions Centre.

To gauge the capabilities of the AI in argument extraction, state-of-art Machine Learning (ML) techniques were utilized. The experiments showed interesting results that support the hypothesis regarding the task's difficulty, which itself likely stems from the subjectivity aspect. In some cases, it is challenging to determine whether a piece of text is a claim, evidence or neither, for instance, a well-known fact or a use of literary devices to make the content more captivating. For example, an epiplexis in the following sentence "*So, if you don't like the new technologies, what solution do you propose?*" [2]. In the presented statement, the writer both challenges the views of those who oppose the new technologies but also implies that it is one of the few or even the only way to go forward. Because of the implied meaning, such a question, especially given the larger context of an article, may be interpreted as a claim. In many cases, the trained ML model rivaled human skill. As an illustration, in the sentence "*And **it is urgent to make it happen**, because **the imbalances will only get worse the longer we wait**.*" [3] the AI identified two components of an argument (excerpts in bold) where the latter elaborates on the former. Of course, the predictions are not always on point, however, in most scenarios the AI's capabilities are not that far off to what an average person can do. Thus, the preliminary results are quite promising, showing potential for practical use in the future.

While the actual proof of concept application did not fit in the timeline of the project, there are a number of possible areas where the automatic extraction of arguments could be utilized to provide a significant value for the users. One of which is debate or writing assistance, where a person could ask the tool to provide compelling arguments on some issue. It could even suggest strong counterarguments to the points made by the opponents. Another interesting use case is with respect to the analysis of actors, essentially profiling the authors of the articles. The tool could summarize the topics that the authors focus on, their opinions and how they reason about them. It could even be extended to organizations that the individuals are a part of, allowing for a larger-scale analysis of groups of individuals with common or opposing narratives that are shared with the public.

To conclude, the project aimed to provide a starting point for structured argument mining research in opinion articles and potentially other types of media such as tweets or forum posts. A dataset of 20 articles was collected as well as annotated and the state-of-art Machine Learning method of automatic argument extraction was examined, showing promising results. Thus, I believe the effort invested in this project will be valuable for future research and, of course, practical applications, enabling more effective communication regarding the importance of sustainability, benefiting both people and the environment.

Referenced articles:

[1] Is 'Hybrid Meat' The Key To Consumer Acceptance Of Cultivated Animal Products?,
https://plantbasednews.org/news/alternative-protein/hybrid-meat-development/
[2] 'Let them eat lentils' won't save us from animal farming – we must embrace meat
substitutes,
https://www.theguardian.com/commentisfree/2023/feb/01/environmentalists-animal-free-meat-livestock-farming
[3] DI Food and Bio: We must utilize alternative proteins if we are to create a CO2 reduction
in the food system,
https://www.altinget.dk/foedevarer/artikel/di-foedevarer-og-bio-vi-skal-udnytte-alternative-proteiner-hvis-vi-skal-skabe-en-co2-reduktion-i-foedevaresystemet