Project Report - Challenge: Argument Mining

Pia Koller, Nina Mandl, Leon Oskui, Mert Türkekul
Koller.Pia@campus.lmu.de, Nina.Mandl@campus.lmu.de,
Leon.Oskui@campus.lmu.de, M.Tuerkekul@campus.lmu.de
Ludwig-Maximilians-Universität — July 9, 2023

Introduction

This report is dedicated to the topic of argument mining and how the topic relates to sustainability. We live in a world where we are confronted with the challenges of sustainability. These challenges bring some problems, which stimulates society to get creative and find solutions to the problem. Such problems are addressed, among other things, in political debates. The discussions are not always enlightening and do not always lead to a good outcome. This is where argument mining can come into action. Argument mining allows us to use automated methods to extract, analyze and understand arguments in text-based sources. The following scenario is an example: We want to discuss whether nuclear power is a solution for the increasing energy demand and climate change. This is a very complex discussion, with many supporters and opponents, each with their own arguments. Furthermore, the topic will probably be taken up several times. With the help of argument mining, the individual arguments of both parties can be filtered out and documented. All arguments are considered afterwards and together a well-founded decision can be made to bring positive changes. Argument mining can be used to prepare for future discussions by reprocessing and revising old arguments. This would generally lead to better documentation and faster solution finding. This is only one of many examples of what argument mining is capable of.

In this project, we set the goal to develop a tool that can be used to apply argument mining to selected texts, for example, to investigate sustainability issues. Sustainability covers a wide range of issues, including environmental problems, social justice, ethical business practices, and more. By applying argument mining, we hope to gain a deeper understanding of the arguments and opinions expressed in public discussions, social media, and other sources on the topic of sustainability.

Problem Statement

In today's information-rich world, discussions on various topics often involve a mixture of pro and con arguments along with a substantial amount of "fluff" - statements lacking factual basis. This abundance of text makes it challenging for individuals to access accurate and relevant arguments, hindering their

ability to gain more in depth insights on specific topics. Moreover, language barriers further limit access to information, as understanding is typically restricted to one's language.

Identifying and extracting arguments from text data poses several challenges. Firstly, the presence of "fluff" statements complicates the process of discerning genuine arguments from irrelevant or unsubstantiated claims. Distinguishing between opinionated statements and well-supported arguments requires sophisticated analysis techniques.

Secondly, text data often lacks a standardized structure, making it difficult to identify and extract individual arguments. The presence of tangential information, digressions, or redundant content further adds to the complexity of the task. Developing algorithms capable of accurately parsing and segmenting text to capture coherent arguments is a significant challenge.

Additionally, language barriers pose a considerable hurdle. Extracting arguments from a text in multiple languages necessitates language-specific processing techniques, such as translation and natural language understanding. Ensuring the accuracy and reliability of argument extraction across diverse languages adds another layer of complexity.

Addressing these challenges is crucial to provide individuals with the means to access well-founded arguments and make informed decisions. Our project aims to tackle some of these hurdles by developing a solution that enables efficient extraction and organization of real arguments from lengthy texts, irrespective of language barriers and the presence of "fluff" statements. By overcoming these obstacles, our solution seeks to enhance the process of acquiring and understanding diverse perspectives on any given topic.

Methodology

In this section, we present the methodology employed to address the challenge of argument mining. We outline the key steps involved in understanding the concept of arguments, developing the argument detection algorithm, merging similar arguments, and considering design and implementation aspects.

Conceptualizing Arguments

To establish a solid foundation for our project, we embarked on a comprehensive exploration of the concept of arguments. Through an extensive literature review, we examined established theories and frameworks in argumentation theory, rhetoric, and logical reasoning. This examination allowed us to gain a deep understanding of the essential elements and characteristics that define an argument within the context of our project. Based on this analysis, we developed a set of criteria for identifying arguments within textual data, taking into account elements such as claim, evidence, reasoning, and logical structure.

Development of Argument Detection Algorithm

Recognizing the need for a tool to overcome our challenges and filter out arguments from various sources like articles, interviews, and papers, we developed a solution to provide everyone with the opportunity to quickly inform themselves to form their own opinions.

We implemented a prototype by leveraging an AI model from Huggingface specialized in argument recognition. The AI model "RoBERTArg"[1] was trained on 25 thousand heterogeneous manually annotated sentences [1] of controversial topics to classify text into one of two labels: NON-ARGUMENT (0) and ARGUMENT (1). The authors of the AI model focus on controversial topics, i.e., topics that include "an obvious polarity to the possible outcomes" and compile a final set of eight controversial topics: abortion, school uniforms, death penalty, marijuana legalization, nuclear energy, cloning, gun control, and minimum wage. In this work, however, the focus was limited to the topic of nuclear energy [2].

To further classify the arguments as pro, con, or neutral, we employed another existing model. The model "bertweet-sentiment-analysis"[3] classifies the arguments as pro, contra, or neutral. The resulting output is a comprehensive table displaying indexed arguments along with their corresponding scores (pro, con, neutral).

Similar Argument Merging Algorithm

To enhance the organization and coherency of the extracted arguments, we developed an algorithm for merging similar arguments together. We adapted our algorithm to effectively capture arguments that span multiple sentences. Initially, the tool searched the text sentence by sentence, determining whether each individual sentence constituted an argument. However, recognizing that arguments often extend beyond a single sentence, we implemented a merging mechanism. Consecutive sentences identified as arguments are now combined into a single argument entity.

The decision to merge consecutive argumentative sentences was based on the observation that the likelihood of consecutive sentences being separate arguments is generally low. Typically, there is supporting "fluff" or explanatory content following an argument. While it is possible for consecutive sentences to contain separate arguments within a text, the current algorithm accounts for this to a reasonable extent.

By grouping similar arguments, we sought to provide a comprehensive and structured representation of the arguments present in the text data.

Design and Implementation Considerations

We created a visual representation of our idea with the prototyping tool Figma1. The application offers multiple options for input, such as uploading a file, inserting an article link, or conducting a general topic search (refer to Figure 1, top left). Users are then presented with an overview of arguments sorted by their ratings, accompanied by graphical representations of the argument-to-fluff ratio and the distribution of ratings.

By following this methodology, we aimed to develop a robust and effective solution for argument mining, enabling users to quickly inform themselves, evaluate arguments, and form their own opinions based on reliable and well-organized information. By combining advanced AI models, efficient categorization algorithms, and an intuitive user interface, our prototype aims to empower individuals to access and comprehend arguments quickly and easily, enabling them to make well-informed decisions and shape their own perspectives.

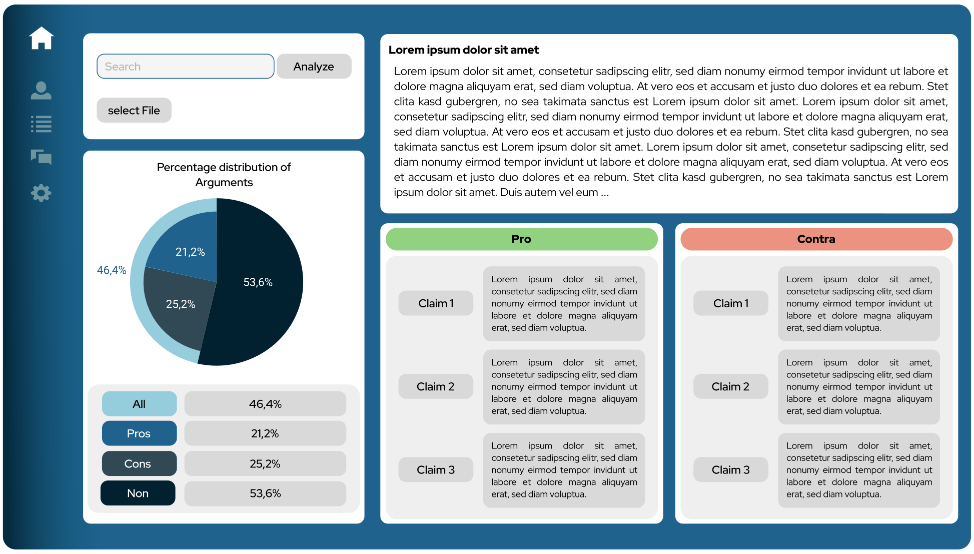

Figure 1: A Figma Prototype Showcasing Our Powerful Tool for Extracting and Analyzing Arguments from Text

Results

Utilizing our Argument Mining prototype, we were able to perform a range of tasks on various files, including reading, analyzing, and subsequently categorizing them. The results generated were meticulously organized in a tabulated format that encapsulated grouped arguments along with their respective evaluations. For ease of access and further analysis, this organized data is stored in a CSV file.

Evaluating the effectiveness and precision of the filtering mechanism used in our prototype is inherently subjective and open to interpretation. Furthermore, as with any innovative development, there's always room for enhancement in the implementation process. Despite these considerations, we confidently assert that the process of categorizing and grouping the arguments has been executed efficiently. However, one aspect that warrants further refinement is the categorization into pro, con, and neutral. This feature, while functional, has not yet reached its fullest potential and requires further fine-tuning to improve its performance.

In addition to manually scrutinizing the arguments, we introduced another layer of examination. We developed a specific prompt (as referenced in Figure 2), which was subjected to testing in both the ChatGPT 3.5 and GPT 4 models. This was conducted with the aim of generating a comparable table of data. The outcomes from this testing phase were highly satisfactory. Particularly noteworthy were the results from GPT 4 and our Argument Mining tool, which produced strikingly similar results. This similarity underscores the successful performance and precision of our Argument Mining tool.

> "Give me a table listing all arguments in the following text. First column is the index, 2nd column is the argument and the 3rd column is if the argument is whether pro or contra. Cite direct and don't summarize the sentences in context. Note that an argument can consist of several connected sentences."

Figure 2: Prompt for ChatGPT 3.5 and GPT 4

However, we observed certain limitations with the ChatGPT 3.5 model during the testing phase. It struggled with effectively grouping the arguments and exhibited a noticeable bias, classifying each argument as positive. This contrasted with the more advanced GPT 4 model, which demonstrated a more balanced approach by successfully identifying and highlighting pro, con, and neutral arguments. This discrepancy underlines the evolution and improvements in language model technology from ChatGPT 3.5 to GPT 4 and our tool.

Another notable advantage of our application lies in its ease of use and convenience. Instead of requiring the user to manually input text into the prompt, our system provides the option to simply upload an external file for analysis. This feature significantly streamlines the process, saving users' time and effort while ensuring comprehensive analysis of the provided data. The capacity to handle and interpret large quantities of data from external files underpins the scalability and user-centric design of our application, making it a high-value tool in the field of argument mining.

Conclusion

As mentioned in the results, the present model was compared with different versions of GPT. This comparison serves among other purposes for the evaluation of our model and especially noticeable was the fact that the model implemented by us showed similar results as the latest version of GPT.
Still, it should be mentioned that the presented product is a prototype. Therefore, it is natural that there is a need for future improvements. In regard to this, it would be important to add a semantic analysis of the arguments to give the argument a stance. Under these conditions, the model is also able to combine similar arguments. Another aspect to improve is the topics on which the model is specialized. Currently, the algorithm can apply argument mining to only 8 topics. For a broader use case, it would be better to use a different model that is not specialized to specific topics.
To conclude the project and incorporate the sustainable context, it can be stated that, despite everything, the current prototype serves as a promising indicator of sustainability. By shedding light on pertinent economic factors, such as in a political debate on environmental issues, our tool enables informed decision-making and effectively prevents unsustainable choices by providing comprehensive information. Hence, our implementation inherently promotes sustainability.

References

[1] Christian Stab, Tristan Miller, Benjamin Schiller, Pranav Rai, and Iryna Gurevych. Cross-topic argument mining from heterogeneous sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3664–3674, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[2] Hugging Face. Robertarg. https://huggingface.co/chkla/roberta-argument last accessed July 2023.

[3] Juan Manuel Pérez, Juan Carlos Giudici, and Franco Luque. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks, 2021.