From Political Debates to Deliberative Democracy: A Roadmap to Assess Semi-Supervised Argument Mining with DISPUTool

Cristian Cardellino, Serena Villata, Elena Cabrio

From Political Debates to Deliberative Democracy: A Roadmap to Assess Semi-Supervised Argument Mining with DISPUTool

Cristian Cardellino, Elena Cabrio, Serena Villata
Université Côte d'Azur, CNRS, Inria, I3S, France
cardellino@i3s.unice.fr, elena.cabrio@unice.fr, villata@i3s.unice.fr

Abstract

Argument mining is a field in natural language processing that studies the automatic extraction of arguments and the classification of their structure from free text. This is a promising research area with numerous applications, like fact-checking, qualitative assessments of online debates, or analysis of legal documents, but with a set of nontrivial challenges to overcome, the main one being the lack of large language resources for the development of this kind of model. In this work, we propose a research path on a subarea of argument mining that has not been explored as much as some other areas: semi-supervised argument mining. We will explore the adaptation of DISPUTool, a model originally trained on data from the US Presidential Elections Political Debates, to deliberative democracy debates within the ORBIS Project for "augmenting participation, co-creation, trust and transparency in deliberative democracy at all scales". Our discussion highlights the technical and non-technical challenges of this task and our planned path of action to overcome them.

Keywords: argument mining, natural language processing, semi-supervised learning

1. Introduction

In the era of the Internet and Social Networks, where lots of people have endless sources of information colliding, while at the same time there are ever-growing applications of generative AI tools like *large language models* (Yenduri et al., 2023) and *text-to-image models* (Zhang et al., 2023), the misuse of these platforms has become a daring challenge both for academia researchers and industry stakeholders.

In the research field of political sciences, for example, there have been well-documented cases of how the use of these tools with the purpose of misinformation can result in manipulations of the decision-making process (Boldyreva, 2018).

As a countermeasure for the abusive use of this platform to the detriment of democracy, there has been an important surge in the research areas of computational social sciences and digital humanities, with the intent of designing frameworks to support social scientists and humanities scholars in their investigations of deliberative discourses that can be used in the process of decision making. Such is the case of the ORBIS[1] project (ORBIS, 2022), which tries to tackle the challenge of citizens increasingly demanding to be engaged in democratic and inclusive discussions.

When dealing with discussions and debates, the issue of analyzing argument structures with natural language processing (NLP) techniques led to a relatively novel research area called Argument(-action) Mining (AM) (Cabrio and Villata, 2018; Lawrence and Reed, 2020). This field deals with the automatic extraction of arguments (e.g., premises, claims, facts, evidence) and the classification of their relations (e.g., attack, support, refute) for analysis of the argumentation structure in texts of different domains. It has applications in medicine, digital humanities, political sciences, qualitative assessments of online debates, fact-checking, etc.

One of the major challenges in AM is the lack of large linguistic resources to train AM models. Unlike some other NLP tasks, where there are large amounts of supervised data or where the supervised data might be easy to generate, AM requires both linguistic experts and domain experts of the field it's being applied to in order to generate high-quality data.

The work by Haddadan et al. (2019) provides a large dataset of political debates from the US presidential election (Haddadan et al., 2019). Moreover, in Goffredo et al. (2023), the authors deliver a modular architecture tool for multi-layer argumentative analysis of political debates.

In this paper, we present a roadmap of our work in progress. We are currently developing an adaptation of DISPUTool's core architecture for model training as a full-fledged Python module (Cardellino, 2024). Using this architecture and the dataset of Haddadan et al. (2019) for training, we plan to experiment in semi-supervised argument mining, assessing the impact of these tools on a dataset in the domain of deliberative democracy.

In the following sections, we describe some of the background work in the area of semi-supervised argument mining, then we describe the set of challenges we face, and finally, describe a path of action for tackling them.

[1] https://orbis-project.eu/

2. Background

When dealing with argument mining (AM), there have been many approaches dealing with supervised scenarios (Cabrio and Villata, 2018). In the last few years, the usage of autoencoding transformers architectures based on BERT (Devlin et al., 2019) have provided for different applications as well (Mayer et al., 2020; Goffredo et al., 2023). However, these approaches generally require annotated linguistic data, which is not always available.

This issue has led to some research in the area of semi-supervised learning for AM. There are different approaches to trying to solve the same issue. One of the first works in semi-supervised AM can be traced back to Habernal and Gurevych (2015) in which the authors explore the usage of large unlabeled resources, the *debate portals*, to produce novel unsupervised features that are added to classic supervised features of AM.

In Wambsganss et al. (2020), the authors tackle argument identification with an iterative approach by using BERT first to identify arguments in unknown corpora and then revising and improving the model based on the analysis of various metrics over the results of the model. They achieve a transfer learning scenario that is corpus-agnostic. This is going to be the base for our future approach.

With the more recent trend of large language models (LLM), new approaches to semi-supervised learning have appeared, mostly in the form of few-shot learning via prompting. The work of Sharma et al. (2023) explores this topic in a multimodality setting. They provide experimentation by comparing the use of LLMs in a few-shot setting with fine-tuned unimodal and multimodal models. The use of LLM-based techniques is also a research area we plan to explore in our scenario.

3. Challenges

Our main challenge to tackle is the complete lack of data in the area of deliberative democracy debates. Deliberative democracy is mostly found in live events like conferences and forums, and many of these events are organized in limited groups and only rely on informal data collection, mostly in the form of meeting notes, minutes, or reports. This is a major issue to tackle since without data to test our models, no matter how good the training data is, it becomes impossible to evaluate a domain adaptation scenario, which is crucial for our work.

There are a couple of events we are surveying in order to see if we can access the data. "The World Forum for Democracy"[2] was an event from last year where many interesting topics were discussed in a deliberation setting. Another event like the "Conference on the Future of Europe"[3] provides a similar setting we are interesting in exploring. Finally, the projects organized by the partners at ORBIS will be another source of data. We are also leveraging the data present in the Bcause platform[4], which offers its own set of challenges, particularly in the use of informal language in the argumentation discourses and the presence of noisy data.

Most of this data, however, is provided in audiovisual formats, which gives us another challenge to deal with. In particular, some events don't provide an English translation of their audio, with speakers stating their arguments in their native language. After finding our sources of data, we need to transcribe such data from audiovisual format to text format so we can process them with our models. Luckily, there are tools available to alleviate this task, and there has been a large improvement in the technologies for speech recognition (Radford et al., 2022). After the transcription, for the case of multilingual data, we plan to use machine translation techniques to ease the process of translating to English since the existing AM models are for English. We know this is not the ideal scenario, but it is a needed step in order to achieve better and more robust systems.

Finally, there is the nontrivial matter of anonymization of the collected data. First and foremost, we need to proceed with the collection of consent from the participants for their data to be used for research purposes and the correct anonymization of the data to avoid potential leakage in accordance with the General Data Protection Regulation of the European Union.

4. Research Rodmap

4.1. Data Collection

We will use the ElecDeb60To16 dataset (Haddadan et al., 2019) as training data. For the development and evaluation of data, we will follow two different processes for collection.

The data for the final evaluation is going to be the more curated one, this mean that it will based on human transcriptions and it will be annotated with human supervision. The dataset will not be large, however, because of the difficulties in data annotation with limited resources. We aim to use it for development and evaluation rather than training. It also has to be as close as possible to the

[2] https://www.coe.int/en/web/world-forum-democracy

[3] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/conference-future-europe_en

[4] https://bcause.app

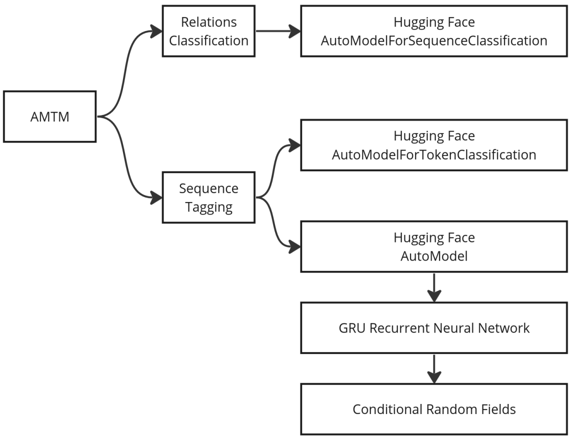

Figure 1: Argumentation Mining Transformers Module Architecture.

distribution of the data we expect to deal with once the final pipeline is ready for deployment.

For the development data, which is required to build the model, we will use semi-supervised annotation techniques. Using the bare version of DISPUTool, we'll annotate the data in a manner that serves as suggestions to aid the annotators. The idea is to alleviate the process of annotation, which is time-consuming and difficult. The data must also be as close to the final data as possible, i.e., it should come from the same or similar distribution of the evaluation data.

4.2. Model Architecture

As established before, our main goal is to assess the semi-supervised argument mining (AM) with DISPUTool 2.0 (Goffredo et al., 2023), trained with ElecDeb60To16, which has a modular architecture and provides a full AM pipeline, both in terms of argument detection and relation classification. We have already implemented the tool as a Python module (Cardellino, 2024) that leverages the power of the Hugging Face library (Wolf et al., 2020), which opens the possibility to access the latest Transformer based architectures for our research.

Figure 1 shows the module architecture. It's basically composed of two modules, one for relation classification based on Hugging Face's models for sequence classification and a sequence tagging module that either uses the Hugging Face token classification module or the same structure based on transformers with recurrent neural networks and conditional random fields of DISPUTool 2.0.

4.3. Approaches for Semi-supervised Argument Mining

Once the development data is collected, the next step is to follow a schema similar to the one proposed by Wambsganss et al. (2020), but extending it further and using it both for argument detection and relation classification. We also plan to explore other autoencoding transformer models, not limiting ourselves to BERT. This will be the first iteration of the process of domain adaptation.

The following iteration is going to add humans to the training loop. Using techniques based on active learning similar to the work of Cardellino et al. (2015), we aim to improve the model development, in particular for cases where it lacks certainty.

Once we reach a plateau for techniques based on domain adaptation, we propose to follow up with few-shot learning like the work of Sharma et al. (2023). Using LLMs fine-tuned on the DISPUTool training data, as well as examples with a distribution similar to that of the development data, we are looking to explore few-shot learning to keep improving the model. With enough fine-tuning and a few shot-learning examples, we should be able to generate large development data for our research.

5. Final Remarks

In this paper, we presented a roadmap to follow within the framework of the ORBIS project. We proposed to adapt an existing tool in the domain of political debates to the more specific domain of deliberative democracy. We stated the challenges we are facing and delineated a path of action to overcome such challenges in the search for a solution to the problem of limited domain-specific data.

6. Bibliographical References

Elena Boldyreva. 2018. Cambridge analytica: Ethics and online manipulation with decision-making process. pages 91–102.

Elena Cabrio and Serena Villata. 2018. Five years of argument mining: a data-driven analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 5427–5433. International Joint Conferences on Artificial Intelligence Organization.

Cristian Cardellino. 2024. Argumentation Mining Transformers Module (AMTM).

Cristian Cardellino, Serena Villata, Laura Alonso Alemany, and Elena Cabrio. 2015. Information extraction with active learning: A case study in legal text. In *Computational Linguistics and Intelligent Text Processing*, pages 483–494, Cham. Springer International Publishing.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Pierpaolo Goffredo, Elena Cabrio, Serena Villata, Shohreh Haddadan, and Jhonatan Torres Sanchez. 2023. Disputool 2.0: A modular architecture for multi-layer argumentative analysis of political debates. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13):16431–16433.

Ivan Habernal and Iryna Gurevych. 2015. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2127–2137, Lisbon, Portugal. Association for Computational Linguistics.

Shohreh Haddadan, Elena Cabrio, and Serena Villata. 2019. Yes, we can! mining arguments in 50 years of US presidential campaign debates. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4684–4690, Florence, Italy. Association for Computational Linguistics.

John Lawrence and Chris Reed. 2020. Argument Mining: A Survey. *Computational Linguistics*, 45(4):765–818.

Tobias Mayer, Elena Cabrio, and Serena Villata. 2020. Transformer-based Argument Mining for Healthcare Applications. In *ECAI 2020 - 24th European Conference on Artificial Intelligence*, Santiago de Compostela / Online, Spain.

ORBIS. 2022. Augmenting participation, co-creation, trust and transparency in deliberative democracy at all scales.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2022. Robust speech recognition via large-scale weak supervision.

Arushi Sharma, Abhibha Gupta, and Maneesh Bilalpur. 2023. Argumentative stance prediction: An exploratory study on multimodality and few-shot learning. In *Proceedings of the 10th Workshop on Argument Mining*, pages 167–174, Singapore. Association for Computational Linguistics.

Thiemo Wambsganss, Nikolaos Molyndris, and Matthias Söllner. 2020. Unlocking transfer learning in argumentation mining: A domain-independent modelling approach. In *Entwicklungen, Chancen und Herausforderungen der Digitalisierung. Band 1: Proceedings der 15. Internationalen Tagung Wirtschaftsinformatik 2020*.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Huggingface's transformers: State-of-the-art natural language processing.

Gokul Yenduri, Ramalingam M, Chemmalar Selvi G, Supriya Y, Gautam Srivastava, Praveen Kumar Reddy Maddikunta, Deepti Raj G, Rutvij H Jhaveri, Prabadevi B, Weizheng Wang, Athanasios V. Vasilakos, and Thippa Reddy Gadekallu. 2023. Generative pre-trained transformer: A comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions.

Chenshuang Zhang, Chaoning Zhang, Mengchun Zhang, and In So Kweon. 2023. Text-to-image diffusion models in generative ai: A survey.

7. Language Resource References

Haddadan, Shohreh and Cabrio, Elena and Villata, Serena. 2019. *USElecDeb60To16 Dataset*. v.01.