

# ACTA

## A Tool for Argumentative Clinical Trial Analysis

**Tobias Mayer , Elena Cabrio and Serena Villata**

Université Côte d’Azur, CNRS, Inria, I3S, France  
 tmayer@i3s.unice.fr, {elena.cabrio, serena.villata}@unice.fr

### Abstract

Argumentative analysis of textual documents of various nature (e.g., persuasive essays, online discussion blogs, scientific articles) allows to detect the main argumentative components (i.e., premises and claims) present in the text and to predict whether these components are connected to each other by argumentative relations (e.g., support and attack), leading to the identification of (possibly complex) argumentative structures. Given the importance of argument-based decision making in medicine, in this demo paper we introduce ACTA, a tool for automating the argumentative analysis of clinical trials. The tool is designed to support doctors and clinicians in identifying the document(s) of interest about a certain disease, and in analyzing the main argumentative content and PICO elements.

### 1 Introduction

Argumentation is the process by which arguments are constructed, compared, evaluated in some respect, in order to establish whether any of them is warranted. In recent years, there has been a growth of interest in the subject from formal and technical perspectives in Computer Science, and a wide use of argumentation technologies in practical applications, ranging from medicine to social media content analysis. The field of artificial argumentation [Atkinson *et al.*, 2017] plays an important role in Artificial Intelligence research. One of the latest advances in artificial argumentation is the so-called *argument(ation) mining* [Lippi and Torroni, 2016; Cabrio and Villata, 2018], whose main goal is to automatically detect the argumentative components in text and to predict the relations holding between them. Argument mining methods have been applied to heterogeneous types of textual documents, e.g., persuasive essays [Stab and Gurevych, 2017], scientific articles [Teufel *et al.*, 2009], Wikipedia articles [Bar-Haim *et al.*, 2017], Web debating platforms [Habernal and Gurevych, 2017], political speeches and debates [Menini *et al.*, 2018]. However, only few approaches [Zabkar *et al.*, 2006; Green, 2014; Mayer *et al.*, 2018a; Mayer *et al.*, 2018b] focused on automatically detecting argumentative structures from textual documents in medicine, such as clinical trials, clinical guidelines, and Electronic Health Records.

In this paper, we present ACTA (Argumentative Clinical Trial Analysis), a tool designed to support clinicians in the analysis of clinical trials. ACTA automatically analyses the textual abstract(s) of clinical trials that the user provides, and it detects in the text the argumentative components (i.e., premises and claims) together with their relations. In addition, we also include the identification of PICO elements in the abstracts.<sup>1</sup> ACTA returns the user with the argumentative structure identified in the selected abstract(s), under the form of a navigable graph whose nodes are the argumentative components. PICO and argumentation elements are highlighted in the textual abstract with different colors.

ACTA employs argument mining methods to identify the argumentative structure of textual clinical trial abstracts. Two stages are crucial in the argument mining framework: (i) the first stage is the identification of arguments within the input natural language text. This step may be further split in two different stages such as the detection of argument components (e.g., claims, premises) and the further identification of their textual boundaries. Many approaches have recently been proposed to address such task, adopting different methods like Support Vector Machines (SVM), Naïve Bayes classifiers, and Neural Networks; (ii) the second stage consists in predicting what are the relations holding between the arguments identified in the first stage. They are used to build the argument graphs, in which the relations connecting the retrieved argumentative components correspond to the edges. Different methods have been employed to address this task, from standard SVMs to Textual Entailment. In this paper, we address the issue of predicting the existence of a link between two argumentative components, without labeling it with a precise relation (e.g., support, attack).

To the best of our knowledge, this is the first tool automatically analyzing textual clinical trials from the argumentative point of view, employing Natural Language Processing (NLP) and Machine Learning methods. ACTA may be seen as the first step of a pipeline ending with evidence-based decision making frameworks in health-care applications, as those illustrated in [Chapman *et al.*, 2019; Hunter and Williams, 2012; Craven *et al.*, 2012; Longo and Hederman, 2013;

<sup>1</sup>PICO is a framework to answer health-care related questions in evidence-based practice. Elements comprise patients/population (P), intervention (I), control/comparison (C) and outcome (O) information.

Qassas *et al.*, 2015]. The demo system of ACTA is available at <http://ns.inria.fr/acta>.

## 2 Argumentative Clinical Trial Analysis

In this Section, we first describe the main features of the ACTA tool, and second, we illustrate the methods we used to set up the classification of argumentative components and the link prediction.

### 2.1 ACTA Main Features

ACTA is a tool designed to ease the work of clinicians in analyzing clinical trials. It goes beyond the basic keyword-based search in clinical trial abstracts, and it empowers the clinician with the ability to retrieve the main claim(s) stated in the trial, as well as the premises (or evidence) linked to this claim. As a result, the clinician does not need to read the whole abstract, but she is provided with a structured “summary” of the abstract under the form of a graph. More precisely, ACTA provides clinicians with the following facilities:

**Search on PubMed.** PubMed<sup>2</sup> is a free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics. Given the importance of this search engine in the health-care domain, we included in ACTA the possibility to search for a (set of) abstract(s) directly on the PubMed catalogue, through their API. When the search results are shown, the user can select one or more abstracts to address the argumentative analysis.

**Argumentative analysis.** As soon as the text is uploaded or the abstract(s) is selected from the PubMed search result list, the user can run the argumentative analysis by pressing the analyse-button. The result is visualized to the user under the form of an argumentative graph where the nodes are the premises and the claims automatically detected in the abstract together with their links. The textual content of the argumentative component is shown, when the user hovers over a node. In addition, the full text of the abstract is shown on the right side of the graph, where premises and claims are highlighted with different colors.

**PICO elements.** In addition, given the importance of the PICO information for the health-care domain, we automatically detect PICO elements in the text of the clinical trial. These elements are highlighted in different colors on the right.

**Download the annotated data.** The result of the argumentative analysis together with all the other information about a study can be downloaded as a json file.

### 2.2 Experimental Setting and Results

For the argumentative component classification and boundary detection, we use a corpus of 500 abstracts of randomized controlled trials on neoplasm treatments annotated with claims, premises and their connections. The relation annotations are later used for the link prediction task. The data is split into two sets: 80% for training and 20% for testing. We

treat the first step of the pipeline as a sequence tagging problem using the BIO tagging scheme, where the goal is to predict for every token if it is the **Beginning**, **Inside** or **Outside** of a component. For this task, the ACTA system relies on a pre-trained bidirectional transformer language model to encode context specific sentence tokens. The token-level representation is input into a recurrent layer, i.e., a bidirectional Gated Recurrent Unit (GRU), with a Conditional Random Field (CRF) layer on top of it. As a transformer architecture, we use the BERT base model [Devlin *et al.*, 2018] with pre-trained weights. We fine-tune the entire model with an Adam optimizer and a learning rate of  $2e-5$  for three epochs. This model achieves a f1-score of 85.2 on the test set.

The same method is applied to train the model for the PICO element extraction. As data, we use the EBM-NLP corpus [Nye *et al.*, 2018] with coarse labels. The model is trained to jointly predict the participant, intervention and outcome candidates for a given input. Dataset splits are the same as in the original paper, with the difference that sentences containing less than 10 WordPiece [Wu *et al.*, 2016] tokens are ignored. Here, the f1-score on the test set is 73.4.

The prediction of the links between argumentative components is treated as a multiple choice problem, similar to the *Situations With Adversarial Generations* (SWAG) task [Zellers *et al.*, 2018], where one has to select the correct target sentence for a sentence-pair from a list of possible candidates. This way we could ensure that we get for each source component at maximum one link to a target component. This is important, since the argumentation graph we try to predict allows one outgoing edge per node at most. For training, we fine-tune the BERT base model with a dense layer for three epochs with an Adam optimizer and a learning rate of  $3e-5$ , resulting in a f1-score of 79.4.

## 3 Concluding Remarks

This paper presents ACTA, a tool for automatically analyzing clinical trial abstracts from the argumentative point of view, highlighting also PICO items. To the best of our knowledge, the combination of these two elements is a unique feature of ACTA. Alas, several activities are ongoing to improve the system. First, we are integrating an active learning module so that we can capture the feedback of clinicians on the classification of argumentative components and link prediction, and employ it to improve the results. Second, we will improve the classification of argumentative components in order to generalize better on different diseases, and improve the way we handle clinical abbreviations. Third, we will also include the labels over the links (attack, support) and take links between different studies into account to fully model relation classification. Finally, we plan to add RobotReviewer [Marshall *et al.*, 2015] to cover the evaluation of the risk of bias of studies.

## Acknowledgements

This work is partly funded by the French government labelled PIA program under its IDEX UCA JEDI project (ANR-15-IDEX-0001).

<sup>2</sup><https://www.ncbi.nlm.nih.gov/pubmed/>

## References

- [Atkinson *et al.*, 2017] Katie Atkinson, Pietro Baroni, Massimiliano Giacomin, Anthony Hunter, Henry Prakken, Chris Reed, Guillermo Ricardo Simari, Matthias Thimm, and Serena Villata. Towards artificial argumentation. *AI Magazine*, 38(3):25–36, 2017.
- [Bar-Haim *et al.*, 2017] Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, and Noam Slonim. Stance classification of context-dependent claims. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017*, pages 251–261, 2017.
- [Cabrio and Villata, 2018] Elena Cabrio and Serena Villata. Five years of argument mining: a data-driven analysis. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 5427–5433, 2018.
- [Chapman *et al.*, 2019] Martin Chapman, Panagiotis Balatsoukas, Mark Ashworth, Vasa Curcin, Nadin Kökciyan, Kai Essers, Isabel Sassoon, Sanjay Modgil, Simon Parsons, and Elizabeth I. Sklar. Computational argumentation-based clinical decision support. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2019*, pages 2345–2347, 2019.
- [Craven *et al.*, 2012] Robert Craven, Francesca Toni, Cristian Cadar, Adrian Hadad, and Matthew Williams. Efficient argumentation for medical decision-making. In *Principles of Knowledge Representation and Reasoning: Proceedings of KR 2012*, 2012.
- [Devlin *et al.*, 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.
- [Green, 2014] Nancy Green. Argumentation for scientific claims in a biomedical research article. In *Proceedings of the Workshop on Frontiers and Connections between Argumentation Theory and NLP*, 2014.
- [Habernal and Gurevych, 2017] Ivan Habernal and Iryna Gurevych. Argumentation mining in user-generated web discourse. *Comput. Linguist.*, 43(1):125–179, 2017.
- [Hunter and Williams, 2012] Anthony Hunter and Matthew Williams. Aggregating evidence about the positive and negative effects of treatments. *Artificial Intelligence in Medicine*, 56(3):173–190, 2012.
- [Lippi and Torroni, 2016] Marco Lippi and Paolo Torroni. Argumentation mining: State of the art and emerging trends. *ACM Trans. Internet Techn.*, 16(2):10, 2016.
- [Longo and Hederman, 2013] Luca Longo and Lucy Hederman. Argumentation theory for decision support in healthcare: A comparison with machine learning. In *Brain and Health Informatics - International Conference, BHI 2013*, pages 168–180, 2013.
- [Marshall *et al.*, 2015] Iain J Marshall, Joël Kuiper, and Byron C Wallace. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *Journal of the American Medical Informatics Association*, 2015.
- [Mayer *et al.*, 2018a] Tobias Mayer, Elena Cabrio, Marco Lippi, Paolo Torroni, and Serena Villata. Argument mining on clinical trials. In *Comp. Models of Argument - Proceedings of COMMA 2018*, pages 137–148, 2018.
- [Mayer *et al.*, 2018b] Tobias Mayer, Elena Cabrio, and Serena Villata. Evidence type classification in randomized controlled trials. In *Proceedings of the 5th Workshop ArgMining@EMNLP 2018*, pages 29–34, 2018.
- [Menini *et al.*, 2018] Stefano Menini, Elena Cabrio, Sara Tonelli, and Serena Villata. Never retreat, never retract: Argumentation analysis for political speeches. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 4889–4896, 2018.
- [Nye *et al.*, 2018] Benjamin Nye, Junyi Jessy Li, Roma Patel, Yinfei Yang, Iain James Marshall, Ani Nenkova, and Byron C. Wallace. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. *CoRR*, abs/1806.04185, 2018.
- [Qassas *et al.*, 2015] Malik Al Qassas, Daniela Fogli, Massimiliano Giacomin, and Giovanni Guida. Analysis of clinical discussions based on argumentation schemes. *Procedia Computer Science*, 64:282–289, 2015.
- [Stab and Gurevych, 2017] Christian Stab and Iryna Gurevych. Parsing argumentation structures in persuasive essays. *Comput. Linguist.*, 43(3):619–659, 2017.
- [Teufel *et al.*, 2009] Simone Teufel, Advait Siddharthan, and Colin Batchelor. Towards domain-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, EMNLP 2009*, pages 1493–1502, 2009.
- [Wu *et al.*, 2016] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Lukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google’s neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144, 2016.
- [Zabkar *et al.*, 2006] Jure Zabkar, Martin Mozina, Jerneja Videcnik, and Ivan Bratko. Argument based machine learning in a medical domain. In *Comp. Models of Argument: Proceedings of COMMA 2006*, pages 59–70, 2006.
- [Zellers *et al.*, 2018] Rowan Zellers, Yonatan Bisk, Roy Schwartz, and Yejin Choi. SWAG: A large-scale adversarial dataset for grounded commonsense inference. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 93–104, 2018.