

Argumentation driven medical research publication analysis and query system

Veronika Solopova

Freie Universität Berlin, Germany
verosolopova@gmail.com

Abstract. This interdisciplinary research aims to tackle relevant and previously unresolved questions of Natural Language Processing, Computational Linguistics, Machine Reasoning, and AI. The system will take a user phrase as an input and give out a synthesized opinion of the scientific community on this matter mined from research publications, with the intermediate steps inter alia consisting of argumentation mining and conflicting views' automated resolution with logical and argumentative structure, the potential presence of bias, extracted as features.

Keywords: Argumentation mining · NLP · Computational Linguistics · logic · bias · public health.

1 Introduction & Research Question

Nowadays, in the era of the search engines, everyone can easily research a question they got interested in or confused about. From Wikipedia to many less-known websites and forums, everyone readily proposes the Internet user an explanation of any concept possible. However, our general Internet behavior and preferences, our geographical position, and the language in which we are searching for the explanation, highly influences what we may find and thus, what we are to believe. Besides, the better part of the websites is written by people without real expertise in the particular field, which becomes especially dangerous for the controversial matters, that can influence our decision-making concerning public health, including diet, alternative versus scientifically proven medicine, and more poignant socio-medical questions. At the same time, the Internet is also a considerable resource of research publications, which represent the opinion backed by the scientific community. However, a common reader who has little to no experience in the domain, usually will not even try to approach such a paper, just because one cannot comprehend its message without the appropriate education. Besides, not all of the papers accessible on PubMed, ResearchGate or Google Scholar, have the same quality, a number of published papers are funded by pharma-companies with their own interests. The system we propose has a multi-layer architecture, which eventually would take the user's statement/question as an input and output the arguments to support or dismiss the statement based on the current scientific community consensus on this matter in

a summarized understandable way. Moreover, it will judge the objective quality of the papers, logical and argumentative structure, and detect potential bias to resolve possible conflicts of interests, presenting both thoughts, for humans to choose.

2 Motivation

This thesis will be concentrating on providing a proof of concept, to demonstrate the feasibility of this task with the currently available techniques. The model cannot be self-sufficient but will be an example of the potential system that can be created based on this study. Neither interface development, nor sufficient covering of the topics will be possible to achieve with the limited resources of a single Ph.D. project. At the same time, the study will undoubtedly contribute to the argumentation mining of the scientific language, which is largely behind the current state-of-art reached for the social media and political speeches. It will also try to tackle nearly unprecedented scientific papers' visual representations parsing and bias detection so that new potential problems can be raised and solved. Finally, it will take a deep insight into the nature of conflict resolution of scientific opinions, including the potentials of the practical embodiment, but also address the ethical questions connected with this task, as well as the very link between the mining and the reasoning, which are important questions no one really asks. We consider that many of these sub-procedures can become a powerful and game-changing tool for the conference publication reviewers, facilitating the process and economizing time. The system can be useful for the research purposes and even prove to be a bottleneck-resolving approach for the meta-analysis studies.

3 Related work

Argumentation is a multi-disciplinary research field, which considers reasoning from the perspective of diverse domains such as logic and philosophy, language, rhetoric, and computer science. Using the theoretical basis of the mentioned above disciplines, automated argumentation mining is actively developing with a number of research projects ongoing on different genres: political debates [1], web discourse including debate portals [2], Twitter and other social media [3,4] and discussion forums [5]. Argumentation mining of the scientific language, including automated theorem proving with the help of shallow semantic embeddings, has been considered by David Fuenmayor and Christoph Benzmlüller in [6] as well as universal meta-logical reasoning [7]. There are also several studies on assessment of argumentative quality and depth [8, 9], argument search engines [10] and retrieval of the counterarguments [11]. Critical analysis of the coherence of the argument is proposed by Scheffler & Stede [12]. Automated bias detection is mainly tackled as a classification task for media and Wikipedia [13, 14]. Conflict detection and resolution are not yet fully understood and thus are usually performed via scoring and assessment based on metadata.

4 Approach

Initial statement analysis. The user’s input, which can be in the form of a statement or a question, is the starting point of the pipeline. It can be a kind of phrase one may enter into the search engine every day, e.d. ‘A good diet’, ‘Do eye exercises actually help?’, ‘Flu/cold?’, ‘Typical cancer symptoms’. The given statement is the only input the system gets, so it should be analyzed carefully to get all the possible information we can, making use of the surface analysis, linguistic analysis, keywords extraction, and semantic parsing, but the system has access to various knowledge sources it uses.

Computational linguistics analysis follows tokenization and shall include Named Entities recognition (for e.x. Monsanto (company), USA (country), Part-of-speech (POS) tagging, syntactic-role parsing, lemmatization (transforming a token to its 0-form which is masculine singular for nouns, adjectives and infinitive for verbal parts of speech. This meta-information is to be added as a tabular description to the tokens.

Keywords extraction is usually done with Machine learning techniques (probably can be done with CRF or Neural networks as in [15]. Semantic parsing is the task of translating linguistic forms into a formal meaning representation on which a machine can act. We plan to realize it as in Broad-Coverage Semantic Parsing as Transduction [16].

Topic correspondence search is part of the information retrieval task and can be done rule-based or with Machine learning if we consider it as a thematic classification task.

Search for similar claims in the suitable documents of the corpus found in step 1. It can be done with Recurrent Neural Networks with different embeddings including the recent BERT[17]. What is also important to keep in mind, is that the majority of studies in the health sector will base their argumentation on the data, collected from the experiments, which is usually presented in the form of graphs, tables and formulas. Thus, ideally the system will have specific parsing mechanisms to mine metadata of the study, analyzing the content inside.

Conflict resolution. Here we come to the most challenging and thus the most scientifically curious part. Obviously, a number of questions do not have a clear answer even in the scientific community, and sometimes it takes several decades to reach a consensus. Consequently, there shall be implemented a conflict resolution, which may give us both standpoints, but also point at the publication with the most impeccable reasoning. Even after filtering all the doubtful publications in the first place, as preprocessing, we can have two or more respectable papers with quite different solutions or views on one problem. This step is of special interest, as it is a complicated reasoning task, even for a human, to decide which one of two acknowledged publications is significantly more convincing. The study is not willing to train a system to decide if a paper is good or bad, but make a pipeline accumulating the objective meta-data, in which case the reasoning of the system can be rule-based.

Evidently, the preprocessing steps, which would initially filter the papers coming into the corpus, are aimed to facilitate the challenge. However, it does not fully

resolve the problem as we cannot simply take, for instance, the newest research, because, new, does not always mean better. Thus, the real basis for preferring any paper shall be found in a profound argumentative and logical chain analysis including information conveyed by the visual representation.

Argumentative structural analysis will include mining of the claims and premises (supports of the argument), quantifying and qualifying them, and searching for the claims that have no supports versus potential implicit premises, which presuppose word and domain knowledge. The logic analysis is done via extracting typical logical structures and sequences, from the previously formalized representation of the text, e.d. ‘if P, then Q’; ‘P only if Q’, ‘Q whenever P’. It also includes verification wherever the text contains one or various types of reasoning: categorical, conditional, disjunctive, conjunctive, etc., and tests if logical chains are formed sound, consistent and coherent. Additionally, there can be made bias detection, which may be done in many ways. One of the potential tests available is the extraction of the names of the companies funding the study and linking them to the companies’ sphere of interests.

The extraction of premises supporting the central claim from the filtered papers is the final step. It implies argumentation mining via semantic parsing or Deep Learning, see e.g. the reconstruction of premises as in [18].

The output of the system can differ depending on the user’s interest: it can answer user’s question, give out the arguments supporting or disclaiming the user’s statement including references to the source, as well as give summarised information about the quality of the paper.

5 Corpus & Evaluation

To built and evaluate the envisioned system, we need the data, research papers describing a range of healthcare questions, which will be issued from the mentioned above PubMed and ResearchGate databases to be added [19]. The annotation shall include argumentative components and relations with rhetorical aspects of scientific writing. The evaluation will be carried out comparing the gold annotations and the predicted ones, based on the precision, recall and F-measure metrics.

References

1. Haddadan, Shohreh & Cabrio, Elena & Villata, Serena & van der Torre, Leendert. (2018). Argumentation Mining in Political Debates. University of Luxembourg, Université Côte d’Azur.
2. Habernal, Ivan & Gurevych, Iryna. (2015). Exploiting Debate Portals for Semi-Supervised Argumentation Mining in User-Generated Web Discourse. 2127-2137. 10.18653/v1/D15-1255.
3. Dusmanu, Mihai & Cabrio, Elena & Villata, Serena. (2017). Argument Mining on Twitter: Arguments, Facts, and Sources. 2317-2322. 10.18653/v1/D17-1245.
4. Lytos, Anastasios & Lagkas, Thomas & Sarigiannidis, Panagiotis & Bontcheva,

- Kalina. (2019). The evolution of argumentation mining: From models to social media and emerging tools. *Information Processing & Management*. 56. 10.1016/j.ipm.2019.102055.
5. Skeppstedt, Maria & Kerren, Andreas & Stede, Manfred. (2019). Finding Reasons for Vaccination Hesitancy: Evaluating Semi-Automatic Coding of Internet Discussion Forums. *Studies in health technology and informatics*. 264. 348-352. 10.3233/SHTI190241.
6. Fuenmayor, David & Benz Müller, Christoph. (2020). Computer-supported Analysis of Arguments in Climate Engineering.
7. Benz Müller, Christoph. (2018). Universal (Meta-) Logical Reasoning: Recent Successes. 10.13140/RG.2.2.11039.61609/2.
8. Wachsmuth, Henning & Al-Khatib, Khalid & Stein, Benno. (2016). Using Argument Mining to Assess the Argumentation Quality of Essays. C16-1158, Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers.
9. Manfred Stede. (2016). Toward assessing depth of argumentation. In Proceedings of COLING 2016. Osaka, Japan.
10. Wachsmuth, Henning & Potthast, Martin & Khatib, Khalid & Ajjour, Yamen & Puschmann, Jana & Qu, Jiani & Dorsch, Jonas & Morari, Viorel & Bevendorff, Janek & Stein, Benno. (2017). Building an Argument Search Engine for the Web. 49-59. 10.18653/v1/W17-5106.
11. Wachsmuth, Henning & Syed, Shahbaz & Stein, Benno. (2018). Retrieval of the Best Counterargument without Prior Topic Knowledge. 241-251. 10.18653/v1/P18-1023.
12. Scheffler Tatjana & Stede, Manfred. (2016). Realizing argumentative coherence relations in German: a contrastive study of newspaper editorials and Twitter posts. In Proceedings of the COMMA Workshop "Foundations of the Language of Argumentation". Potsdam, Germany.
13. Herzig, Livnat & Nunes, Alex & Snir, Batia. (2011). An Annotation Scheme for Automated Bias Detection in Wikipedia.. 47-55.
14. Baraniak, Katarzyna & Sydow, Marcin. (2018). News articles similarity for automatic media bias detection in Polish news portals. 21-24. 10.15439/2018F359.
15. Pay, Tayfun. (2016). Totally automated keyword extraction. 3859-3863.
16. Zhang, Sheng & Ma, Xutai & Duh, Kevin & Durme, Benjamin. (2019). Broad-Coverage Semantic Parsing as Transduction. 3777-3789. 10.18653/v1/D19-1392.
17. Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
18. Winkels, Radboud & Douw, Jochem & Veldhoen, Sara. (2014). State of the ART: an Argument Reconstruction Tool. 10.13140/2.1.1012.4162.
19. Lauscher, Anne & Glavaš, Goran & Ponzetto, Simone. (2018). An Argument-Annotated Corpus of Scientific Publications. 40-46. 10.18653/v1/W18-5206.