Automated argumentation mining to the rescue? Envisioning argumentation and decision-making support for debates in open online collaboration communities

Jodi Schneider

▶ To cite this version:

HAL Id: hal-01076322
https://hal.archives-ouvertes.fr/hal-01076322

Submitted on 21 Oct 2014

Automated argumentation mining to the rescue? Envisioning argumentation and decision-making support for debates in open online collaboration communities

Jodi Schneider[*]
INRIA Sophia Antipolis, France
`jodi.schneider@inria.fr`

Abstract

Argumentation mining, a relatively new area of discourse analysis, involves automatically identifying and structuring arguments. Following a basic introduction to argumentation, we describe a new possible domain for argumentation mining: debates in open online collaboration communities. Based on our experience with manual annotation of arguments in debates, we envision argumentation mining as the basis for three kinds of support tools, for authoring more persuasive arguments, finding weaknesses in others' arguments, and summarizing a debate's overall conclusions.

1 Introduction

Argumentation mining, a relatively new area of discourse analysis, involves automatically identifying and structuring arguments. Following a basic introduction to argumentation, we describe online debates as a future application area for argumentation mining, describing how we have manually identified and structured argumentation, and how we envision argumentation mining being applied to support these debates in the future.

1.1 What is an argument

Informally, an argument is a communication presenting reasons for accepting a conclusion. Unlike proofs that lead step-by-step from premises with logical justifications for a conclusion, arguments are non-monotonic and can be disproven. Arguments may use various approaches including generalization, analogy, inference, and prediction.
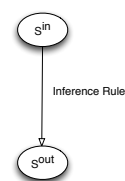
Figure 1: The simplest possible argument.

The simplest possible argument connects two Statements by means of an Inference Rule (Figure 1). Inference Rules are functions that input one or more Statements (the premises) and return one or more Statements (the conclusions).

1.2 More complex arguments

Far more complex arguments can be formed. Arbitrary numbers of arguments can be joined into a larger and more complex argument. Useful terminology is introduced by (Wyner et al., 2008), who reserve the term *argument* to refer to the simplest kind: non-decomposable arguments. They distinguish *cases* which support a single conclusion (see Figure 2) from *debates* which argue for and against a single conclusion.
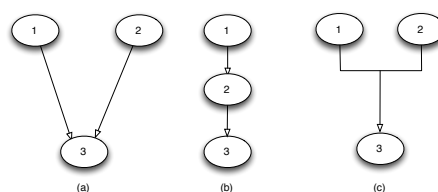
Figure 2: Cases support a single conclusion. Cases may (a) use multiple, independent premises to support a single conclusion; (b) draw an intermediate conclusion, and use it as an additional premise in order to support a final conclusion; or (c) require two linked premises (both required as input to the inference rule) to support a conclusion.

Figure 3 shows a simple debate, where two arguments *attack* one another. There are three ways
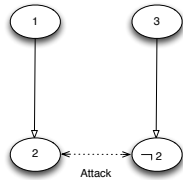
Figure 3: *Debates* argue for and against a single conclusion. This kind of attack is called a *rebuttal*.

of attacking an argument: attacking a premise (known as *undermining*), attacking a conclusion (known as *rebutting*), and attacking an inference (known as *undercutting*), following (Prakken, 2010).[1]

1.3 Inference Rules

Argumentation schemes, e.g. (Walton et al., 2008) are one way of expressing Inference Rules. These are *patterns* for arguing which are stated abstractly: to use an argumentation scheme, it must be instantiated with details. To indicate possible flaws in reasoning, associated with each scheme there are critical questions pointing to the possible counterarguments.

We next introduce an example from our own work, where automated argumentation mining could be used.

2 Rationale-based debate in open online communities

One place where argumentation mining could be applied is in rationale-based debate in open online communities. The Web has enabled large-scale collaboration, even among people who may never meet face-to-face. A large number of participants present their views and reasoning to make decisions for open, online collaborative software and knowledge development in Mozilla, Wikipedia, OpenStreetMap, etc. In these groups, asynchronous textual debates are the basis for decision making. Participants argue for decisions based on rationales, since the *reasons* for opinions, rather than majority votes or aggregate sentiment, justify decisions. Thus large-scale decision support in these communities should make evident not just the overall tendency of the group (as in opinion mining) but rather the arguments made, focusing

especially on the rationales, or reasons given for a preferred outcome.

In our work, we have analyzed a corpus of debates, to understand how the English-language version of Wikipedia makes decisions about which articles to include and exclude from the encyclopedia. We used two approaches to argumentation theory to annotate asynchronous messages in each debate, in iterative multiparty annotation experiments (Schneider, 2014).

2.1 Analysis using argumentation schemes

First, we used Walton's argumentation schemes (outlined in Ch. 9 of (Walton et al., 2008)) in order to annotate the arguments, focusing on the internal reasoning of each message. First one person (this author) annotated all the arguments found in the corpus against Walton's 60 schemes, finding 1213 arguments in 741 messages (Schneider et al., 2013). Then, we focused on the subset of 14 argumentation schemes that appeared more than 2% of the time, with iterative, multiparty annotation. There was a sharp divide between the two most prevalent argument types–*Argument from Evidence to Hypothesis* (19%) and *Argument from Rules* (17%)–and the remaining 12 types that appeared from 2-4% of the time.

Besides these patterns, we found statistically significant differences between how experts and novices in the community argued in our corpus of debates. Experts were more likely to use *Argument from Precedent*, while novices (who had little experience in the debates and in the wider Wikipedia community) were more likely to use several argumentation schemes that the community viewed as less sound bases for decision making.[2] These included *Argumentation from Values*, *Argumentation from Cause to Effect*, and *Argument from Analogy*.

2.2 Analysis using factors analysis

Second, we used a very different approach, based on factors analysis (Ashley, 1991) and dimensions theory (Bench-Capon and Rissland, 2001), which

---

[1]Rebut and undercut are drawn from the well-known work of (Pollock, 1994); Prakken credits undermining to (Vreeswijk, 1993) and (Elvang-Gøransson et al., 1993).

[2]Our analysis of acceptability of arguments drew from community documentation and took community responses to messages into account. For instance, *Argumentation from Values* might be countered by a messages saying "Whether you personally like an article or its subject, is totally irrelevant." (This exchange appeared in our corpus in fact `http://en.wikipedia.org/wiki/Wikipedia: Articles_for_deletion/Log/2011_January_ 29`.)

have most commonly been used in case-based reasoning. We iteratively derived four factors important in the discussions: Notability, Sources, Maintenance, and Bias (Schneider et al., 2012). This was an easier annotation task, with stronger inter-annotator agreement than for Walton's argumentation schemes: factors analysis had Cohen's kappa (Cohen, 1960) of .64-.82 depending on the factor (Schneider et al., 2012), versus .48 for Walton's argumentation schemes (Schneider et al., 2013)). Factors provide a good way to organize the debate; filtering discussions based on each factor can show the rationale topic by topic, which supported decision making in a pilot user-based evaluation (Schneider, 2014).

We can also identify the misunderstandings that newcomers have about which factors are important, and about what kind of support is necessary to justify claims about whether a factor holds. When an article is unacceptable because it lacks reliable sources, it is not enough to counter that *someone will publish about this website when it gets out of beta testing*.[3] This newcomer's argument fails to convincingly establish that there are reliable sources (because for Wikipedia, a reliable source should be published, independent, and subject to full editorial control), and may make things worse because it suggests that the sources are not independent. Rather, a convincing counterargument would explicitly address how the most relevant criteria are met.

3 Envisioned applications of argumentation mining

The manual annotations described above, of argumentation schemes and of factors, suggest several possibilities for automation. Scalable processes for analyzing messages are needed since Wikipedia has roughly 500 debates each week about deleting borderline articles. Argumentation mining could be the basis for several support tools, helping participants write more persuasive arguments, find weaknesses in others' arguments, and summarize the overall conclusions of the debate.

First consider how we might give participants feedback about their arguments. From our research, we know which argumentation schemes are viewed as acceptable and persuasive within the community. If real-time algorithms could identify the argumentation schemes used in the main argument, authors could be given personalized feedback even before their message is posted to the discussion. When the argumentation scheme used in a draft message is not generally accepted, the author could be warned that their message might not be persuasive, and given personalized suggestions. Thus debate participants might be nudged into writing more persuasive arguments.

Next consider how we could help participants find weaknesses in others' arguments. Automatically listing critical questions might benefit the discussion. Critical questions point out the possible weaknesses of an argument, based on the argumentation scheme pattern it uses. Listing these questions in concrete and contextualized form (drawing on the premises, inference rules, and conclusions to instantiate and contextualize them) would encourage participants to consider the possible flaws in reasoning and might prompt participants to request answers within the debate. In the authoring process, supplying the critical questions associated with argumentation schemes might also help the author (who could consider elaborating before submitting a message).

Finally, we could envision argumentation mining being used to summarize the debate. Macro-argumentation, such as the factors analysis described above, would be a natural choice for summarization, as it has already proven useful for filtering discussions. A more reasoning-intensive approach would be to calculate consistent outcomes (Wyner and van Engers, 2010), if debates can be easily formalized.

3.1 Challenges for argumentation mining

In previous work, argumentation schemes have been classified in constrained domains, especially in legal argumentation (Mochales and Moens, 2011) and by using (Feng, 2010; Feng and Hirst, 2011) the Araucaria corpus (Katzav et al., 2004).[4]

Each of our envisioned applications of argumentation has certain requirements. Automatically detecting the argumentation schemes used in a message could be used for supporting authoring and finding weaknesses of arguments, which focus on the interior of each message. In order to ask the

---

[3]This is a real argument from a newcomer from our corpus, slightly reworded for clarity.

[4]Further work is needed on argument scheme prevalence, which seems to vary by domain. Only 3 of Feng's 5 'most common argumentation schemes' appear in the top 14 most common schemes in our corpus, excluding *Argument from Example* and *Argument from Cause to Effect*.

appropriate critical questions, the premises, conclusions, and inference rules would first need to be detected. To get at the point of each message, the macro-level argumentation (for instance using factors and dimensions) would be useful for summarizing the debate, especially if we record rationales.

Another challenge is to create scaleable architectures for real-time or batch reprocessing of argumentation mining on the Web. In our scenarios above, support for authoring arguments would require real-time feedback (i.e. within minutes). Slower batch processing would be useful for the two other scenarios (support in challenging arguments with critical questions; support for summarizing debates) since Wikipedia's debates are generally open for 7 days.

3.2 Related scenarios

This is a single use case, but it represents a wide array of related ones. Open source and open knowledge projects are full of decision making discussions available widely in textual form. Rhetorical studies of them so far take place on a qualitative, discursive level. Examples include dissent and rhetorical devices in bug reporting (Ko and Chilana, 2011) and how Python listservs select enhancement proposals (Barcellini et al., 2005). Interestingly, the role of a participant in the Python community is related to the kinds of message they quote (Syntheses, Disagreements, Proposals, or Agreements), and Syntheses and Disagreements are the most quoted. The organizational relevance of these open decision making discussions in collaborative communities makes them a promising target for support, and argumentation mining technology is an appropriate tool to deploy towards that end.

4   Conclusions

This paper detailed how automated argumentation mining could be leveraged to support open online communities in making decisions through online debates about rationale. We first gave a basic overview of argumentation structures, describing arguments as consisting of Statements, Inference Rules, and (possibly) Attacks. Then we described our own work on manual identification of argumentation schemes in Wikipedia information quality debates. We envisioned three kinds support tools that could be developed from auto-

mated argumentation mining in the future, for authoring more persuasive arguments, finding weaknesses in others' arguments, and summarizing a debate's overall conclusions. Open online communities are a wide area of application where argumentation mining could help participants reason collectively.

References

Kevin D Ashley. 1991. *Modeling Legal Arguments: Reasoning with Cases and Hypotheticals*. MIT Press.

Flore Barcellini, Françoise Détienne, Jean-Marie Burkhardt, and Warren Sack. 2005. A study of online discussions in an open-source software community. In *Communities and Technologies 2005*, pages 301–320. Springer.

Trevor J M Bench-Capon and Edwina L Rissland. 2001. Back to the future: Dimensions revisited. In *Proceedings of JURIX 2001*, pages 41–52.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Morten Elvang-Gøransson, Paul J Krause, and John Fox. 1993. Acceptability of arguments as 'logical uncertainty'. In *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 85–90. Springer.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies–Volume 1*, pages 987–996.

Vanessa Wei Feng. 2010. Classifying arguments by scheme. Master's thesis, University of Toronto.

Joel Katzav, Chris Reed, and Glenn Rowe. 2004. Argument Research Corpus. In *Proceedings of the 2003 Conference on Practical Applications in Language and Computers*, pages 229–239. Peter Lang.

Andrew J Ko and Parmit K Chilana. 2011. Design, discussion, and dissent in open bug reports. In *Proceedings of the 2011 iConference*, pages 106–113.

Raquel Mochales and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law*, 19(1):1–22.

John L Pollock. 1994. Justification and defeat. *Artificial Intelligence*, 67(2):377–407.

Henry Prakken. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124.

Jodi Schneider, Alexandre Passant, and Stefan Decker. 2012. Deletion discussions in Wikipedia: Decision factors and outcomes. In *Proceedings of the International Symposium on Wikis and Open Collaboration*, pages 17:1–17:10.

Jodi Schneider, Krystian Samp, Alexandre Passant, and Stefan Decker. 2013. Arguments about deletion: How experience improves the acceptability of arguments in ad-hoc online task groups. In *Proceedings of the ACM conference on Computer Supported Cooperative Work*, pages 1069–1080.

Jodi Schneider. 2014. *Identifying, Annotating, and Filtering Arguments and Opinions in Open Collaboration Systems*. Ph.D. dissertation, Digital Enterprise Research Institute, National University of Ireland, Galway. Corpus and supplementary material also available online at `http://purl.org/jsphd`.

Gerard Vreeswijk. 1993. *Studies in Defeasible Argumentation*. Ph.D. dissertation, Free University Amsterdam.

Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge.

Adam Wyner and Tom van Engers. 2010. Towards web-based mass argumentation in natural language. In *Proceedings of Knowledge Engineering and Knowledge Management 2010 Poster and Demo Track*.

Adam Z Wyner, Trevor J Bench-Capon, and Katie Atkinson. 2008. Three senses of "Argument". In *Computable Models of the Law: Languages, Dialogues, Games, Ontologies*, pages 146–161. Springer-Verlag.