A Pilot Study in Using Argumentation Frameworks for Online Debates

Federico CERUTTI [a], Alexis PALMER [b], Ariel ROSENFELD [c], Jan ŠNAJDER [d] and Francesca TONI [e]

[a] *Cardiff University, U.K.*
[b] *Universität Heidelberg, Germany*
[c] *Bar-Ilan University, Israel*
[d] *University of Zagreb, Croatia*
[e] *Imperial College London, U.K.*

Abstract. We describe a pilot study in using argumentation frameworks obtained from an online debate to evaluate positions expressed in the debate. This pilot study aims at exploring the richness of Computational Argumentation methods and techniques for evaluating arguments to reason with the output of Argument Mining. It uses a hand-generated graphical representation of the debate as an intermediate representation from which argumentation frameworks can be extracted, but richer than any existing argumentation framework. The intermediate representation can provide insights for benchmark sets derived from online debates.

Keywords. Argumentation frameworks comparison; Benchmarks; Argument Mining

1. Introduction

Computational Argumentation (CompArg) is a branch of AI aiming at providing computational models of argumentation; see [6], [7], and [21] for overviews. In its simplest form, CompArg amounts to characterising and determining (dialectically) acceptable sets of arguments in any given *Abstract Argumentation* Framework [14], consisting simply of a set of abstract entities (the *arguments*) and a binary relation of *attack* between arguments. Several other forms of CompArg have been proposed and deployed in applications, including the Argumentation Framework with Recursive Attacks (AFRA) [3,4], allowing attacks to be in turn the object of other attacks, and *Quantitative Argumentation Debate* (QuAD) Frameworks [5,19], allowing graded (numerical) acceptability statuses of arguments [5,19,8].

This paper describes a pilot study in comparing those different frameworks, building on top of an Argument Mining (ArgMin) exercise. ArgMin is an emerging field aiming to automatically extract argumentation structures from natural language texts; see [17,18, 16] for overviews. To this end, it heavily relies on Natural Language Processing (NLP) to detect the argumentative discourse structure in text and recognize the components of an argument and relations between them.

Despite a large theoretical investigation on the semantic *intertranslatability* of frameworks [10], to our knowledge there have been no attempts to compare different

frameworks with respect to real-world tasks. With this aim, we considered a pipeline approach, where the output of ArgMin provides an input to tools developed within CompArg to determine the dialectical acceptability and/or strength of opinions in debates.

The starting point of the experiment was an excerpt from an online for/against debate taken from `www.createdebate.com`. The excerpt is given in Table 1.[1] We then:

1. mapped the debate onto a hand-annotated graphical representation, identifying annotations dynamically as demanded by the features of the debate and the opinions expressed therein; this resulted in a rich annotation scheme with five types of nodes and six types of edges;
2. mapped the hand-annotated graphical representation onto an Abstract Argumentation Framework and determined the dialectical acceptability of opinions in the debate by determining the grounded labelling [2] of the arguments in the framework;
3. mapped the hand-annotated graphical representation onto a QuAD Framework and used the Arg&Dec tool ([1], `www.arganddec.com`) to determine the dialectical strength of opinions in the debate, as well as to rank the two answers (yes/no) to the debated question;
4. mapped the hand-annotated representation onto an Argumentation Framework with Recursive Attacks and determined the dialectical acceptability of opinions by using the grounded extension [4];
5. compared the results obtained with those frameworks.

The pilot study raises a number of questions, from both the ArgMin and CompArg perspectives. In particular, for CompArg:

- whether our hand-annotated graphical representation can be used as a tool for producing cross-framework benchmarks;
- whether the Argumentation Frameworks and tools considered are sufficiently general to serve as a target for reasoning automatically with debates;
- whether other existing Argumentation Frameworks and tools may be more suitable for the task at hand.

The paper is organised as follows. Section 2 first presents in full the debate used as a starting point for our experiment, then continues with the hand-annotated graphical representation of the debate. We additionally discuss the relationships between user-generated annotations to the dialogue and those from expert annotators. Section 3 shows the mapping onto Abstract Argumentation, Section 4 onto QuAD, and Section 5 onto AFRA. In Section 6 we conclude.

2. A Graphical Analysis of the Input Debate

We identified statements in the dialogue as well as the relationships between them via the means of a graph-based representation. Although this representation has been influenced by other works, notably the Argument Interchange Format (AIF) [13,20], and Inference

[1] For the full debate see `http://goo.gl/DZuRdg`

Debate question: **Should contraception be covered by health insurance?**
#1 Intangible *Noes because that's not something you need.*
#2 sweetspice16 (disputes #1) *You probably shouldn't make that blanket statement, without any qualifiers or exceptions. For many women, birth control pills are very important and are necessary to daily life.*
#3 Cartman (disputes #2) *What about Viagra, should that be covered by health insurance?*
#4 Sitara (disputes #1) *Women have the right to choose what to do with their bodies.*
#5 ThePlague (disputes #4) *It is true that women have the right to choose what they wish to do with their bodies, but they have absolutely no power to force insurance companies to pay for them. That should be left up to the insurance company, and not the woman.*
#6 sweetspice16 (disputes #5) *Oh please. Contraception doesn't have to be to prevent pregnancy either. I nearly went broke paying for birth control pills and I was on them because of severe issues. But insurance doesn't have to pay for it even then. Men don't need erections but Viagra is covered in case a patient has other issues. Birth control should be covered too: no matter what, just in case.*
#7 ThePlague (disputes #6) *That is a much more logical argument that the user Sitara. Allow me to continue. I agree with you. I do not think Viagra should be covered by insurance though and thus do not believe that contraception should be provided by insurance companies. It should not be covered due to it's initial purpose, to prevent pregnancy, hence the name "birth control".*
#8 Sitara (disputes #7) *Wrong. I am presenting a very logical argument.*
#9 ThePlague (disputes #8) *You cannot follow the purpose of the debate. She used a conparative argument. Viagra is covered yet contraception is not? That is a much more solid argument since it follows the premise of this debate. Your argument is over women's right to choose contraception. This debate doesn't call for that.*
#10 Sitara (disputes #9) *I have presented a logical argument. I told you why contraception should and will be covered by insurance, but you choose to ignore logic. Do stop wasting my time.*
#11 ThePlague (disputes #10) *Contraception should be required because it is a women's right to choose? If a murderer wishes to purchase a weapon to use for mass slaughter will you favor his decision as well since he has the right to choose what he wants? No. The company has the right to deny service to him and thus can do the same with contraception. You cannot favor the liberty of women without favoring the liberty of a business.*
#12 Sitara (disputes #11) *Logical fallacy. Contraception is not comparable to murder.*

Table 1. An excerpt from an online debate.

Anchoring Theory [11], it has been developed in an ad-hoc fashion driven by a linguistic view point. A comparative study with other approaches is left for future work.

In particular, we considered five types of nodes:

- *question nodes*;
- *answer nodes*;
- *standard statements*;
- *partial statements*—statements with missing premises or conclusions, i.e., enthymemes;
- *distractor statements*—statements that are dialectically irrelevant, albeit on topic.

We linked the nodes using six types of edges, each taking one of several different possible values, as follows:

- *answer-to-question*, from one answer node to a question node (directed edges);
- *standard-explicit*, from one standard statement node to another, or to a distractor node, a partial statement node, or an answer node (directed edges), with possible values **attack/support/neither**;
- *standard-implicit*, from one partial statement to any statement or answer node (directed edges), with possible values **attack/support**;
- *meta*, from any statement to any statement (directed edges), with possible values **attack/support**;
- *node-to-edge*, from standard statements to edges (directed), with possible values **attack/support**;
- *expansion*, amongst any statements (undirected edges).

This analysis resulted in the graph shown in Fig. 1, with Q denoting the question node, and Y and N denoting the answer nodes. Moreover, we label each node in the graph with the identifier of the statements made in the debate (e.g., "#2" in Table 1). Some identifiers in Fig. 1 have a superscript (i.e., 2) to indicate that they actually represent multiple (i.e., two) statements.

For example, we made the following mapping choices in deriving the graph:

- #4 is a partial statement as it lacks an explicit conclusion;
- #3 is a distractor statement as it is a sort of distraction from the main point of the debate, although still "on topic", and could be interpreted as intended to promote conflict; edges onto distractor statements are neither attacks nor supports;
- #2 and #6 form an expansion statement because #6 fills in some of the details omitted in #2;
- #12 criticises #11 at the dialectical (meta) level as well as the content (standard) level;
- #8 criticises the attack by #7 on #4.

Comments on the Annotations. Once a debate has started in the system, users may posit arguments in the form of short textual posts as seen in Table 1. However, as shown in the expert annotation presented in Fig. 1, some of these posts contain more than a single argument, which poses the challenge of splitting posts into atomic arguments.

Furthermore, each user of the debate platform is required to explicitly define how her posts correspond and relate to the existing posts that were already presented in the debate. Specifically, the user is required to choose whether her post **supports**, **disputes**,

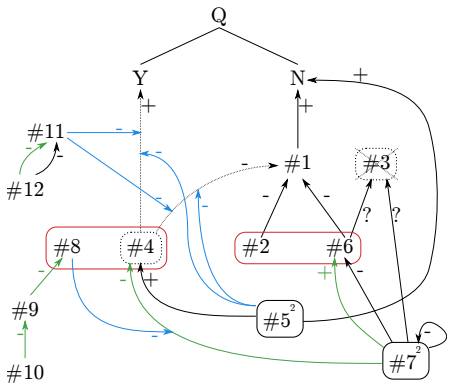

Figure 1. Output of the initial graphical analysis. Dotted nodes represent partial statements, crossed dotted nodes represent distractor statements, straight lines represent answer-to-question edges, solid black arrows represent standard-explicit edges, dotted black arrows represent standard-implicit edges, green arrows represent meta edges, blue arrows represent node-to-edge edges, and red rounded boxes represent expansion edges. A $+$ (resp. $-$) next to an arrow identifies support (resp. dispute); a ? is used when the nature of the relation between the two statements is unclear.

or **clarifies** an existing post. We consider these as non-expert annotations of the relations between the presented posts.

In the debate on which we focus, all posts were annotated as **dispute** posts by their authors. Namely, all posts were annotated as disputing other posts that had already been presented in the debate. However, in our post factum annotation in Fig. 1, we show that sometimes it is the support and clarification relations that were actually intended.

In some cases, dispute annotations can be interpreted as attack annotations. For example, post #2 was designated as disputing post #1 by its author and indeed argument #2 attacks argument #1 (see Fig. 2). Yet, this is not always the case. For example, post #6 was designated as disputing post #5, however we did not find any significant relation between these posts.

Overall, it is our opinion that having non-expert annotations generated by debaters can be useful as a rough starting point for expert or automated annotation of the relation between arguments. Nevertheless, one needs to keep in mind that these non-expert annotations are biased and imperfect.

3. From the Graphical Analysis to an Abstract Argumentation Framework

We mapped the hand-annotated graphical representation given in Fig. 1 onto an Abstract Argumentation Framework in order to determine the dialectical acceptability of opinions in the debate. To this aim, we had to identify the two main components of an Abstract Argumentation Framework, namely the set of arguments and the set of attacks. In fact, an Abstract Argumentation Framework [14] is composed by a set of arguments whose nature is left unspecified, and by a binary relation of attacks among them. Therefore, an Abstract Argumentation Framework can be represented as a directed graph, where nodes identify arguments, and edges attacks.

Since the notion of *argument* is now overloaded with different meanings, in this and in the following section, *argument* stands for *formal abstract argument*, i.e., an element

of a mathematical theory of computational argumentation. We refer to the pieces of texts considered in the annotation process as *statements*.

3.1. Identification of Arguments

To compute the dialectical acceptability of opinions in the debate, it was necessary both to include the two possible outcomes of the dialogue—i.e., whether a player would answer Yes (Y) or No (N) to the question—and to link arguments to the statements put forward in the dialogue. In particular, we needed to identify *atomic statements*—as each player might put forward multiple atomic statements in a single claim. We then aggregated atomic statements into arguments.

3.1.1. Identification of Atomic Statements

The first step is to identify the atomic statements in the dialogue. According to Fig. 1, nodes #5 and #7 contain two statements each:

- #5a: *It is true that women have the right to choose what they wish to do with their bodies,...*;
- #5b: *...but they have absolutely no power to force insurance companies to pay for them. That should be left up to the insurance company, and not the woman.*;
- #7a: *That is a much more logical argument that the user Sitara. Allow me to continue. I agree with you.*;
- #7b: *I do not think Viagra should be covered by insurance though and thus do not believe that contraception should be provided by insurance companies. It should not be covered due to it's initial purpose, to prevent pregnancy, hence the name "birth control*.

The other statements require no further analysis and thus are treated as atomic.

3.1.2. Aggregation of Atomic Statements into Arguments

We then aggregated atomic statements into arguments by exploiting both implicit and explicit support links, as well as expansion links. Therefore, #2 and #6 together form the argument #2#6, and similarly #4 together with #5a and #8. However, expansion and support play different roles: an expansion should be interpreted as a single argument that spans multiple atomic statements. Support should rather be seen as a combination of two sub-arguments. We chose to also represent sub-arguments in the Abstract Argumentation Framework, and thus #4#8 should be considered as an additional argument, as well as #5a alone.

3.2. Identification of Attacks

To simplify the discussion, we assumed that the arguments Y and N are mutually exclusive, and thus attacking each other. Therefore an implicit or explicit support to a positive answer to the question (respectively a negative answer to the question) is transformed into an attack to the negative answer (respectively the positive answer). Since both #5b and #1 support (cf. Fig. 1) the negative answer to the question, they now both attack the

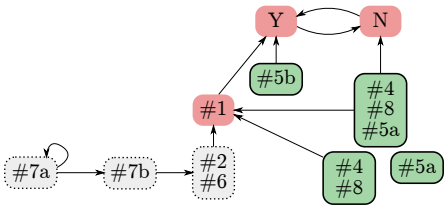

Figure 2. Abstract Argumentation Framework derived from the analysis in Fig. 1. This figure also shows the grounded labelling for this Framework: green nodes (solid, strong border) are IN, red nodes (no border) are OUT, and gray nodes (dotted border) are UNDEC.

Y argument. Similarly, #4 supports a positive answer to the question and thus it attacks the N argument.

Moreover, we also considered attacks derived from the attacking links, either explicit or implicit, depicted in Fig. 1. Therefore, the argument #2#6 attacks #1, and similarly the argument #4#8 (and clearly its super-argument comprising #4, #8 and #5a) attacks #1.

Finally, #7b attacks the argument #2#6, while #7a is a self-defeating argument that also undermines #7b.

3.3. Relevance to the Dialogue and Filtering

The analysis depicted in Fig. 1 requires a language much richer than just abstract arguments and attacks. The links marked with a question mark as well as those denoting meta-information are rather complicated to represent in the abstract formalism. In this pilot study we chose to ignore them instead of enforcing a specific semantics that—in our opinion—is still unclear.

Similarly, Fig. 1 includes edges pointing to other edges, potentially implying other sorts of meta-information. Although there are proposals for encompassing recursive attacks on Abstract Argumentation Frameworks [4], for the sake of this work we chose once again to rely only on Dung's original proposal which does not allow such cases—i.e., attacks are only between arguments.

Consequently, #3, #9, and #10 become unconnected arguments. Similarly, #12 attacks #11, but together they are detached from the rest of the graph. Since they cannot have any effect whatsoever on the dialectical acceptability of opinions for this dialogue, in particular they cannot influence the acceptability of arguments Y or N, we chose to filter them out from the final Abstract Argumentation Framework depicted in Fig. 2.

3.4. Dialectical Acceptability of Arguments

Once the Abstract Argumentation Framework depicted in Fig. 2 is obtained, we can evaluate the dialectical acceptability of each argument by identifying *positions*—i.e., sets of arguments—that together stand against critiques and form a coherent point of view. In [14] several criteria are proposed for such a task, and each criterion identifies a specific position, or *extension* using Dung's terminology, given an Abstract Argumentation Framework. Those criteria can be in terms of labellings: an exhaustive discussion on this topic is beyond the scope of this paper, interested readers are referred to [2]. In short, in a *complete* labelling, an argument is labelled IN if all its attackers are OUT (which clearly

includes the case that the argument is unattacked), OUT if at least one of its attackers is labelled IN, and UNDEC otherwise. The set of IN arguments in a complete labelling is in one-to-one correspondence to a complete extension [2]: therefore, the unique complete labelling, which is depicted in Fig. 2, identifies also the grounded extension (which is the minimal w.r.t. set inclusion complete extension) as well as the unique preferred extension (which are maximal w.r.t. set inclusion complete extensions) [14] of this Abstract Argumentation Framework.

Both Y and N are OUT as a combined effect of #5b and the argument comprising #4, #8, and #5a. Although inconclusive, it allows participants in the dialogue to strategically focus their attention. Indeed, let us assume that participants are supporting the Yes answer, then they should focus on arguing against #5b as it is the only argument undermining the Y argument.

4. From the Graphical Analysis to a QuAD Framework

4.1. Identification of Arguments and Attacks

As a next step, we mapped the hand-annotated graphical representation given in Fig. 1 onto a QuAD Framework [5], and input this into the Arg&Dec tool[2] to determine the dialectical strength of opinions in the debate, as well as to rank the two answers (Yes/No) to the debated question.

We followed the same approach described in Section 3.1 to identify arguments. Unlike Abstract Argumentation Frameworks, though, QuAD Frameworks allow both attack and support relationships between arguments to be represented explicitly, by assigning "types" to arguments (as *pros* or *cons* or *answers*). Thus, arguments #4#8 and #5a can keep their separate identities in the resulting QuAd Framework, and #5a is no longer "isolated". Moreover, QuAD Frameworks, when visualised as graphs, are acyclic, with the result that neither the mutual attack between the Y and N arguments nor the self-attack by argument #7a can be represented directly in the resulting QuAD Framework (see Fig. 3, where pros arguments are indicated with '+', cons arguments are indicated as '-' and answer arguments are indicated by a blue light-bulb/mushroom). Note that, since arguments have a single "type" in QuAD Frameworks, if an argument simultaneously attacks one argument and supports another, it (and all its descendants, if any) needs to be duplicated, as in the case of argument #4#8 in Fig. 3.

Note also that converting the original graphical analysis in Fig. 1 to the QuAD Framework in Fig. 3 required simplifications similar to those for converting to Abstract Argumentation (Section 3.3).

4.2. Dialectical Strength of Arguments

In QuAD Frameworks, arguments are assigned a dialectical strength, from which, in particular, a ranking amongst answer arguments is determined. Note that ranking answers amounts to seeing them as "incompatible"; thus the lack of mutual attacks between answers is not a genuine limitation of QuAD Frameworks.

[2] `www.arganddec.com`

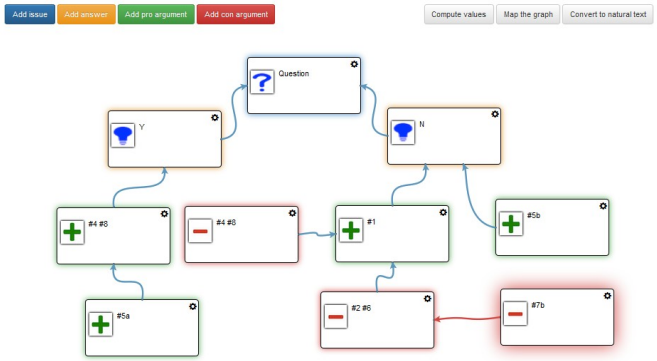

Figure 3. QuAD Framework derived from the analysis depicted in Fig. 1, as depicted in Arg&Dec (`www.arganddec.com`).

In order to determine the strength of arguments in QuAD Frameworks, they need to have a base score to start with (seen as an intrinsic strength, prior to any debate about the arguments). Note that the self-attacking argument #7a in the original Fig. 1 can be thought of as having a base score of 0, because of the self-attack, amounting to its computed strength being also 0 (by using, for example, the methods for computing strength in [5,19]). This renders the argument *ineffective* [5] and justifies its exclusion from the QuAD Framework in Fig. 3.

We experiment with two different policies for assigning base scores to the arguments included in Fig. 3, leading to different rankings of the answers using the method in [5]:

1. All arguments have a medium strength (0.5) to start with; this choice results in Yes being ranked higher than No (with computed strengths, respectively, 0.875 and 0.796875);
2. All arguments have a medium strength (0.5) to start with except

   - Argument #2#6, with a base score close to the maximum allowed (1), by virtue of the supporting meta edge from argument #7;
   - Argument #4#8, with a base score close to the minimum allowed (0), by virtue of the attacking meta edge from argument #7.

   Choosing base scores 0.9 for #2#6 and 0.1 for #4#8 results in No being ranked higher than Yes (with computed strengths, respectively, of 0.811875 and 0.775).

Thus, the use of base scores in QuAD Frameworks can accommodate information (e.g., meta-edges) playing no role in Abstract Argumentation Frameworks. Morever, the use of dialectical strength instead of dialectical acceptability of arguments can help better discriminate amongst arguments, but is highly sensitive to the choice of underlying base score. Indeed, it is clear that the choice of base scores influences the final outcome from the system.

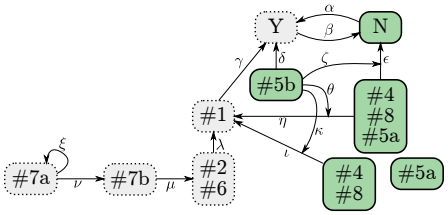

Figure 4. AFRA derived from the analysis depicted in Fig. 1. In green (solid, strong border), the grounded extension (restricted to the set of arguments) for this Framework.

5. From the Graphical Analysis to an AFRA

5.1. Identification of Arguments and Attacks

As the final step, we mapped the hand-annotated graphical representation (Fig. 1) onto AFRA [4], and input this into the Aspartix [15] tool[3] to determine its grounded extension.

We followed the same approach described in Section 3.1 to identify arguments.[4] Unlike Abstract Argumentation Frameworks, though, AFRA allows attacks to be in turn the object of other attacks. Therefore, we are now able to represent the attacks from #5^2 to the attacks between #4#8#5a and #1, (similarly between #4#8 and #1), and between #4#8#5a and N. The resulting framework is depicted in Fig. 4. Please note that attacks to supports such as the one from #5^2 against the support from #4 to Y (Fig. 1) becomes an attack on the attack from #5a#8#4 against N (Fig. 2, see discussion in Section 3.2).

5.2. Dialectical Strength of Arguments

The semantic notions of AFRA are derived from those that apply for Dung's Abstract Argumentation Framework [14]. The main difference is that attacks will also participate as active actors and thus they can also be part of a semantics extension. In particular, the grounded extension of the AFRA depicted in Fig. 4 is $\{$N, #5b, #4#8#5a, #4#8, #5a, $\alpha, \delta, \zeta, \theta, \kappa\}$. Thus, in this representation, the No answer is accepted.

Fig. 4 also depicts the restriction of the grounded extension to the set of arguments only. The ζ attack, in particular, is pivotal in defending the argument N from the attack it received from #4#8#5a, which is instead effective when using only the Dung's framework (Section 3.1).

6. Conclusion

In this paper we discuss a pilot study for comparing different argumentation frameworks on the basis of the same annotation resulting from an analysis of an online debate. The analysis suggests that the information captured by the original annotation scheme (Fig. 1) is much richer than what can be represented in some of the current state-of-the-art frame-

[3] `https://www.dbai.tuwien.ac.at/proj/argumentation/systempage/`

[4] Although AFRA allows to represent more interactions than Dung's AF, e.g., #11 attacking the support from #4 to Y, we chose to consider the same set of arguments identified in Section 3 in order to facilitate the comparison among the different formalisms.

works and tools. We also lack a ground truth (for assessing which position debated is strongest) to assess which tool is better equipped for the task of analysing the specific dialogue we considered in this pilot study.

In the case at hand, increasing the elements of the original annotation schema included in the formal analysis, i.e., the case of AFRA, leads to a less undecided situation w.r.t. the outcome of the dialogue. Moreover, the use of graded semantics as in QuAD allows a much more fine-grained analysis and shows how initial assumptions on the base score of each argument might have a sensible effect on the outcome of the dialogue.

Apart from highlighting differences between Abstract Argumentation, QuAD, and AFRA, this pilot study shows how the proposed annotation scheme (Fig. 1) seems well equipped to represent the complexity of online debates, and that it could be used to produce a set of benchmarks for a variety of frameworks. In fact, most—if not all—of the process described in Sections 3, 4, and 5 can be easily automatised. The foremost issues are determining the arguments (cf. Section 3.1.1), and the relationships among them, especially considering that non-expert annotations are of little help (cf. Section 2).

This pilot study may also help in linking the two research areas of Argument Mining (ArgMin) and of Computational Argumentation (CompArg). In particular, we showed that the output of a potential ArgMin process—namely, the graphical analysis in Fig. 1—may become the input to tools developed in the CompArg community for determining the dialectical acceptability or strength of opinions in debates. Moreover, this mapping may provide valuable feedback to debaters, for example, to inform strategies regarding which aspects to focus on in order to modify the outcome of debates, or to make decisions based on debates. Concretely, we mapped a naturally-occurring multi-party debate from a debate website onto a hand-annotated graphical representation, and then: (a) onto an Abstract Argumentation Framework to determine the dialectical acceptability of opinions [14]; (b) onto a QuAD Framework [5] to determine the dialectical strength of opinions using the Arg&Dec tool [1]; and (c) onto an AFRA Framework [4] to encompass more elements of the original analysis (Fig. 1).

Future work will include evaluating other frameworks proposed in CompArg, e.g., ADF [9], or GRAPPA [10], for representing debates at the level of detail required by the annotations described in Fig. 1. Also, as the investigation of human perception and behavior in argumentative interactions is becoming more prominent in argumentation research [12,22], future work will also include a more thorough investigation of how non-expert annotations made by human debaters can be used by automatic tools.

This pilot study raises a number of questions also for the ArgMin community, while at the same time shedding some light on the applicability of the approach taken. For instance it would be interesting to study whether any of the existing NLP methods and tools could be deployed to support the automatic generation of the initial graphical representation and annotation scheme. Moreover, it would be interesting to study other debates to ascertain the generality or otherwise of the annotation scheme we identified.

Acknowledgements

The input debate was suggested by Adam Wyner and Ivan Habernal, as part of the Dagstuhl seminar on "Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments". We also thank Ivan Habernal for helpful feedback during the preliminary graphical analysis of the debate described in Section 2.

References

[1] M. Aurisicchio, P. Baroni, D. Pellegrini, and F. Toni. Comparing and integrating argumentation-based with matrix-based decision support in arg&dec. In *Theory and Applications of Formal Argumentation - Third International Workshop, TAFA 2015, Buenos Aires, Argentina, July 25-26, 2015, Revised Selected Papers*, pages 1–20, 2015.

[2] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.

[3] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. Encompassing Attacks to Attacks in Abstract Argumentation Frameworks. In C Sossai and G Chemello, editors, *ECSQARU 2009*, pages 2–7. LLNAI, Springer-Verlag, 2009.

[4] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. AFRA: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning (Special Issue Tenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty - ECSQARU 2009)*, 52(1):19–37, 2011.

[5] P. Baroni, M. Romano, F. Toni, M. Aurisicchio, and G. Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49, 2015. Special issue: Applications of logical approaches to argumentation.

[6] T. J. M. Bench-Capon and P. E. Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171(10-15):619–641, 2007.

[7] P. Besnard and A. Hunter. *Elements of Argumentation*. The MIT Press, 2008.

[8] E. Bonzon, J. Delobelle, S. Konieczny, and N. Maudet. A Comparative Study of Ranking-based Semantics for Abstract Argumentation. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)*, pages 914–920, 2016.

[9] G. Brewka, S. Ellmauthaler, H. Strass, J. P. Wallner, and S. Woltran. Abstract dialectical frameworks revisited. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence IJCAI 2013*, pages 803–809. AAAI Press, aug 2013.

[10] G. Brewka and S. Woltran. GRAPPA: A Semantical Framework for Graph-Based Argument Processing. In *21st European Conference on Artificial Intelligence2*, pages 153—-158, 2014.

[11] K. Budzynska and C. Reed. Whence inference? Technical report, University of Dundee, 2011.

[12] F. Cerutti, N. Tintarev, and N. Oren. Formal arguments, preferences, and natural language interfaces to humans: an empirical evaluation. In *ECAI*, pages 207–212, 2014.

[13] C. I. Chesnevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. R. Simari, M. South, G. A. W. Vreeswijk, and S. Willmot. Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04):293, December 2006.

[14] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321 – 357, 1995.

[15] U Egly, S A Gaggl, and S Woltran. Answer-Set Programming Encodings for Argumentation Frameworks. Technical Report DBAI-TR-2008-62, Technische Universität Wien, 2008.

[16] M. Lippi and P. Torroni. Argument Mining: A Machine Learning Perspective. In *The 2015 International Workshop on Theory and Applications of Formal Argument*, 2015.

[17] M-F. Moens. Argumentation mining: Where are we now, where do we want to be and how do we get there? In *Post-proceedings of the forum for information retrieval evaluation (FIRE 2013)*, 2014.

[18] A. Peldszus and M. Stede. From argument diagrams to argumentation mining in texts: A survey. *Int. J. Cogn. Inform. Nat. Intell.*, 7(1):1–31, January 2013.

[19] A. Rago, F. Toni, M. Aurisicchio, and P. Baroni. Discontinuity-free decision support with quantitative argumentation debates. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa*, 2016.

[20] I. Rahwan, B. Banihashemi, C. Reed, D. Walton, and S. Abdallah. Representing and classifying arguments on the Semantic Web. *The Knowledge Engineering Review*, 26(04):487–511, November 2011.

[21] I. Rahwan and G. R. Simari. *Argumentation in Artificial Intelligence*. Springer, 2009.

[22] A. Rosenfeld and S. Kraus. Providing arguments in discussions based on the prediction of human argumentative behavior. In *AAAI*, pages 1320–1327, 2015.