

Argument Analytics

John LAWRENCE^a, Rory DUTHIE^a, Katarzyna BUDZYNSKA^{a,b}, and Chris REED^a

^a*Centre for Argument Technology, University of Dundee, UK*

^b*Institute of Philosophy and Sociology, Polish Academy of Sciences, Poland*

Abstract. Rapid growth in the area of argument mining has resulted in an ever-increasing volume of analysed argument data. Being able to store information about arguments people make in favour or against different opinions, decisions and actions is a highly valuable resource, yet extremely challenging for sense-making. How, for example, can an analyst quickly check whether in a corpus of citizen dialogue people tend to rather agree or disagree with new policies proposed by the department of transportation; how can she get an insight into the interactions typical of this specific dialogical context; how can the general public easily see which presidential candidate is currently winning the debate by being able to successfully defend his arguments? In this paper, we propose Argument Analytics – a suite of techniques which provide interpretation of, and insight into, large-scale argument data for both specialist and general audiences.

Keywords. Argument Interchange Format, corpus resources, argument visualisation

1. Introduction

Over the last decade, a lot of effort has been made to provide various tools for argument analysis, evaluation and visualisation (*cf.* [6]). Systems cover various domains of applications such as analysis and visualisation of reasoning structure (e.g. Carneades [5], Rationale [13]), visualisation of debates [10] or evaluation of argument acceptability [4, 12]. The Argument Interchange Format (AIF) [2], supports exchange and data reuse between these tools, and is currently the only mechanism that handles linguistic, structural and abstract facets of argumentation. Though AIF representations are scalable, analytical tools struggle at scale: for example, an OVA analysis of a single, 45-minute episode of the BBC Radio 4 program *Moral Maze*¹ contains over 200 statements with a similar number of connections between them²; comparable large-scale analyses can be created with Rationale, Carneades and so on. Whilst such maps enable us to follow chains of reasoning, and answer questions about the relationships between individual points, they fall short of providing clear insight into the nature of what is taking place within the argument. With rapid growth of corpora of analysed arguments resulting in part from increased coherence in analysis techniques, and in part from improving results from argument mining systems, the task of making sense out of argument data is becoming increasingly important. Something more than mere visualisation is required.

¹<http://www.bbc.co.uk/programmes/b006qk11>

²<http://www.arg-tech.org/AIFdb/argview/789>

Argument Analytics provides a suite of techniques for analysing AIF data, with components ranging from the detailed statistics required for discourse analysis or argument mining, to infographic-style representations, offering insights in a way that is accessible to a general audience. The extensible set of modules currently comprises: simple statistical data (Section 3), which provides both an overview of the argument structure and frequencies of patterns such as argumentation schemes; comparative data (Section 4) providing a range of measures describing the similarity of two analyses; dialogical data (Section 5) highlighting the behaviour of participants of the dialogue; and real-time data (Section 6) allowing for the graphical representation of a developing over time argument structure. Together these analytics open an avenue to giving feedback on live debates, producing summaries of deliberative democracy, mapping citizen science, and more.

2. Foundations

The Argument Analytics platform is designed specifically for making sense out of argument data represented according to the Argument Interchange Format(AIF) [2] such as the data stored in the AIFdb³ database [7]. The Social Layer [11] is used to enrich this data, providing details on participants such as biographies. AIFdb Corpora enables Argument Analytics to display the interpretations of data, whether on a single AIF argument map (stored in AIFdb as a NodeSet), or a large corpus containing hundreds or thousands of such AIF representations.

The AIF was developed as a means of describing argument networks that would provide a flexible, yet semantically rich, specification of argumentation structures. These networks are comprised of seven types of node:

Node Type	Description
I	propositional information contained in an argument, such as a conclusion, premise, data etc.
L	subset of I-nodes referring to propositional reports specifically about discourse events
RA	application of a scheme of reasoning or inference
CA	application of a scheme of conflict
MA	application of a scheme of rephrasing
YA	application of a scheme of illocution describing communicative intentions which speakers use to introduce propositional contents
TA	application of a scheme of interaction or protocol describing relations between locutions

3. Simple Statistics

The simple statistics modules allows an analyst to quickly make sense of a large amount of annotated argument data. Although these calculations are straightforward and relatively easy to automate, they nevertheless provide interesting insights into the data.

³<http://www.aifdb.org>

The **overview** page shows a range of statistics, offering a rapidly digested summary of the overall argumentative structure. The number of Information nodes provides an indication of the overall size of the analysis. The average number of words per Information Node illustrates the complexity of the ideas presented, and how succinctly they are expressed. The numbers of inference (RA) and conflict (CA) nodes give a suggestion as to the nature of the dialogue, which is further expanded by showing the ratios of RA to CA (capturing how diverse are the perspectives in the debate) and RA to I (how dense the argumentation is).

The **Pattern Count** modules expand on the overview to give detailed statistics suitable for more in-depth argument and discourse analysis. They provide the frequencies of commonly occurring patterns, split into two categories. Firstly, argumentative and illocutionary patterns which describe both the nature of the interactions, for example levels of agreement and disagreement, and the way in which participants have expressed themselves and interacted with each other, such as how frequently a participant questions the statements of others compared to how frequently they assert their own views. The second category, dialogical patterns, illustrates the flow of the discourse and gives an indication of any dialogical rules, either explicit or implicit, to which the participants are conforming. Such dialogical patterns are also useful, for instance, to show cross-cultural differences in dialogue, or differences in the formality and setting of dialogues.

4. Comparative Statistics

The comparative statistics modules [3] allow for the validation of both manual and automatic analysis for argument mining (*cf.* [8,9]). Such calculations enable comparison between two manual analyses to determine the efficacy of annotation guidelines via inter-annotator agreement, or the comparison of results from automatic techniques to a manually created gold standard. The examples given in this section refer to two human annotators, but in each case the same calculations could be applied with one of these being an annotation produced by an automatic system.

There are a number of considerations that must be taken into account when calculating agreement or results, such as what effect a differing segmentation of the original text, in two separate annotations, may have on the assignment of inference and conflict in an argument structure. To account for this, the agreement and results calculations were split into smaller sub-calculations covering segmentation similarity, propositional contents (inference and conflict) and dialogical contents (locutions). Calculating agreement for segmentation of argumentative units is a challenging task [14]. The modular architecture of Argument Analytics allows for a range of measures to be displayed, and currently differences are accounted for using various segmentation similarity algorithms, which give an overall normalised score for the similarity. Propositional contents are compared by separating nodes from the text and instead using the Levenshtein distance for the matching of nodes. Dialogical contents are compared in the same way with word ordering added to the Levenshtein distance for node matching and with the addition of added calculations for the intricacies of dialogue (see [3] for an in-depth description of the comparative statistics module).

5. Dialogically Oriented Statistics

For those argument analyses where there is a dialogue taking place between multiple participants, a range of dialogically oriented, analytics modules are able to provide insights into the dynamics of the discourse, and make these complex interactions accessible to a general audience. There is growing demand to present complex argumentative structures to a broad audience in ways which are both intuitive and interactive. Whilst there is some progress towards this goal, for example, the Election Debate Visualisation Project [10], many of these approaches rely on custom, genre-specific interfaces for both the elicitation and display of argumentative structure. Dialogically oriented, analytics modules make use of both the locution details stored in AIFdb, as well as the participant details provided by the Argument Web social layer.

Each of the modules in this section are illustrated using data from an episode of the BBC Radio 4 program *Moral Maze*⁴. These examples show how such graphical displays of information can take the technical details captured in the argumentative structure of a complex debate, and present them in ways which are easily processed by a general audience.

5.1. Structural Statistics

The structural statistics modules extract particular facets of the argumentative structure in order to display data such as who is speaking most, which pairs of participants are interacting most and who is making the most well supported arguments. As such, they provide a greater insight into the argumentative structure than that which is afforded by looking at a simple argument map of the same data.

Participation: For each participant, the number of locutions they have made is counted and represented in a bar chart. This provides an easy way of identifying which participants were most, and least, dominant within a dialogue. An example can be seen in Figure 1, which shows that Jan Macvarish was the most active participant in this dialogue with twenty-three locutions, whereas Matthew Taylor was least active with only one locution made.

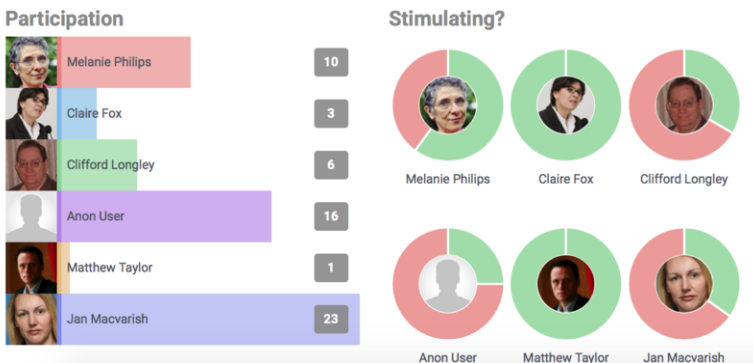


Figure 1. Graphical representations of the relative involvement of each participant in a dialogue, and how stimulating the points made by each participant are.

⁴<http://www.bbc.co.uk/programmes/b006qk11>

Stimulating: A point of debate is stimulating if it receives responses, either to agree or disagree. From the analysed argument structure, we count the number of locutions which each participant has made that have at least one response, and those which have been ignored by the other participants. The example in Figure 1 shows that whilst Claire Fox has only made three locutions, they have all been responded to in some way, whereas, of the six locutions made by Clifford Longley, only two received any attention from the other participants.

Chord Diagram: The chord diagram shows the interaction between participants. A chord diagram is a graphical method of displaying the inter-relationships between data in a matrix. The data is arranged radially around a circle with the relationships between the points drawn as arcs connecting the data together. In this case, the arcs represent interaction between participants, with the width of the arc at each end representing the number of locutions made by that participant to which the connected participant has responded. Viewing the interactions in this way makes it easy to identify, for example, cliques. An example chord diagram can be seen in Figure 2. Clicking on a specific participant emphasises their connections with other participants. For example, with Melanie Phillips selected (as shown on the right of the figure), we can see that the majority of her interactions were with Jan Macvarish, reflecting the fact that, for a period of the dialogue, Melanie was questioning Jan.

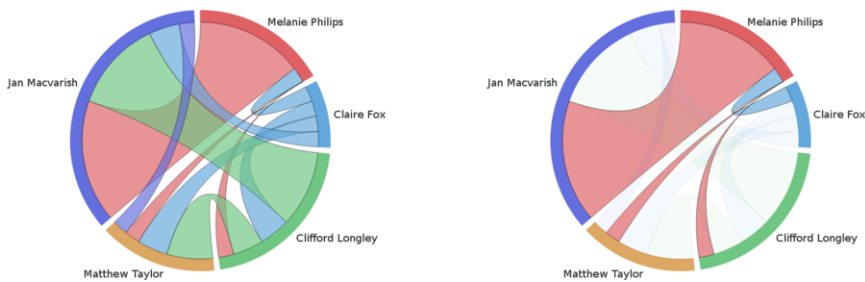


Figure 2. Chord diagrams showing the frequency of interactions between participants. The diagram on the right shows Melanie Phillips selected, highlighting just those interactions in which she is involved.

Verbosity: Similar to the average number of words per I-node presented in the overview, verbosity shows a comparison of the average length of locutions made by each participant. By comparing in this way, we are able to see not just the overall complexity of the ideas expressed, but also how prolix or concise each participant is in presenting their ideas.

5.2. Temporal Statistics

Temporal statistics use the time-stamping of locutions provided by AIFdb to show how the state of a dialogue has altered as it has progressed. These statistics provide clues, not easily discernible from an argument map, as to when individual participants have been most involved in the dialogue, when conflict has arisen, and changes in topic that have occurred as the dialogue progresses.

Turn Structure: Using the timestamping of locutions provided by AIFdb, a graphical representation of the turn structure in a dialogue is created. This visualisation pro-

vides a quick overview of when each participant has been most active, suggesting details of any pre-defined turn-taking rules. The example shown in Figure 3 reflects the turn structure in a *Moral Maze* episode. As the episode begins, each of the four regular panelists speak briefly about the topic being discussed. A guest witness is then introduced, and, after providing their own views on the topic, are then questioned by first one of the panelists and then by a second.

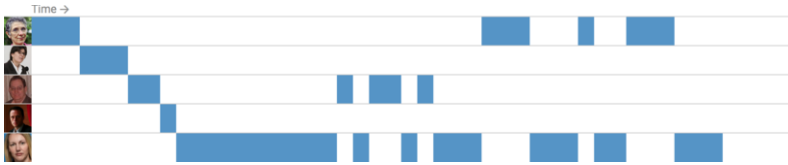


Figure 3. Graphical representation of the turn structure in a dialogue, highlighting the way in which each participant introduces themselves, followed by direct interactions between two pairs of participants.

5.3. Semantics-based Statistics

Semantics-based analytics use Dung-style semantics to determine the acceptability of a participant's arguments. An AIF graph is translated into ASPIC⁺ then, using TOAST, a Dung-style abstract argumentation framework is derived and evaluated.

Defended: The defended points in a dialogue, are those where conflicting points have been made, but these conflicting points have, in turn been attacked. It is easy in a broad ranging and complex dialogue for points to be made which are not challenged either due to going unnoticed, or being simply dismissed. By looking at those points which are challenged and then later defended we gain an insight into both the validity of a point, and how crucial it is to the argument which a participant is making.

Sway: Where one participant has more acceptable arguments than another, the former is said to carry more sway. This value is calculated for each participant, and displayed as the relative balance in sway between each pair of the most commonly interacting participants. This can, to some extent, be viewed as who is winning in a debate; best supporting their own points and best attacking the points made by the other participants in the dialogue.

6. Real-time Statistics

Many of the modules used in Argument Analytics have the ability to not only display data on a fixed, pre-analysed argument structure, but to update in real-time as the structure evolves. This functionality has been used, for example, in a tool developed for the *Built Environment for Social inclusion through the Digital Economy (BESiDE)* project⁵, to facilitate round table discussions between architects working on the design of care environments, and the various stakeholders involved in the design process.

As the discussion is taking place, the audio is recorded and an analyst uses a custom-designed interface to segment the dialogue when either the topic or the speaker changes.

⁵<http://beside.ac.uk/>

A simple dialogue protocol is used, allowing participants to make moves of various types (e.g. asking questions, agreeing with another participant, and offering their own opinion), and relating to a set of pre-defined topics relevant to the design project.

Throughout the discussion, the dialogue overview shown in Figure 4 is displayed for all participants to see. This overview includes a transcript of the dialogue on the right hand side, and analytics modules displaying how much each participant has spoken, and which topics have been discussed on the left. In testing these interfaces, it is interesting to see that they serve not only an informative function, but actually impact the dynamics of the dialogue. When a participant can see that they are talking more than everyone else, they tend to let others speak more. When someone hasn't spoken yet, the other participants notice this, and make an effort to direct questions at them. And, when one topic has been less explored than the others, there is a noticeable shift towards that area in both the questions asked and the points raised.

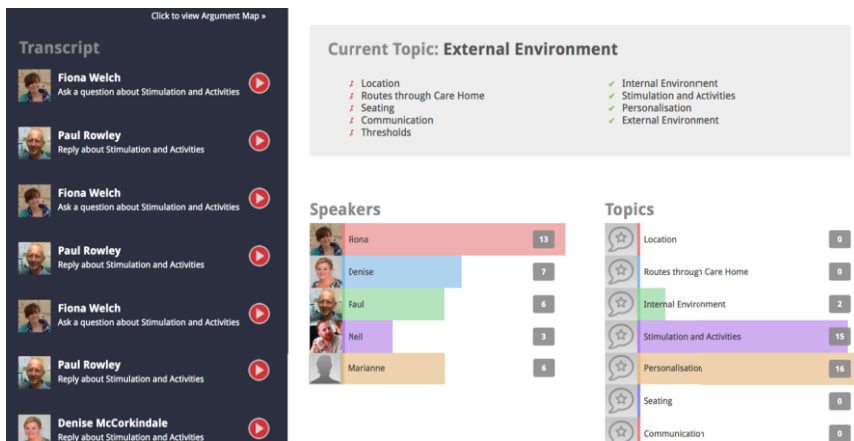


Figure 4. Real-time Argument Analytics highlighting the involvement of individual participants and the topics discussed.

This ability for the argumentative and dialogical structure to, not only represent the outcome of a discussion, but to inform the participants and help ensure that all areas are fully explored has wide ranging potential applications. The current limitation to providing this kind of interface more widely is the ability to perform real-time analysis, but as tools, such as the AnalysisWall [1] which has been used to analyse several hour-long radio programmes in real time, improve, and automatic argument mining techniques develop, it is easy to imagine such a live display accompanying activities such as debates, meetings and media coverage.

7. Conclusions

The Argument Analytics suite provides a comprehensive range of analytic tools from the detailed statistics required for discourse analysis, to graphic visual representations making the same data accessible to a general audience. The existing modules which we have described offer solutions to a broad range of potential user groups, including those involved in argument analysis and critical discourse analysis, those working on argument

mining applications, people performing political or social studies, and members of the general public who wish to get a greater understanding of the issues and dynamics of a complex debate.

There are a range of existing tools which provide argument analysis and visualisation capabilities, but, by using the ability of AIFdb to translate the output from many of these tools into an Argument Interchange Format compliant representation, Argument Analytics allows their output to be displayed in a far broader range of ways and with a broader range of potential applications than any one of these tools currently provides.

Acknowledgments

We would like to acknowledge that the work reported in this paper has been supported in part by EPSRC in the UK under grants EP/M506497/1, EP/N014871/1 and EP/K037293/1.

References

- [1] Floris Bex, John Lawrence, Mark Snaith, and Chris Reed. Implementing the Argument Web. *Communications of the ACM*, 56(10):66–73, Oct 2013.
- [2] Carlos Chesñevar, Sanjay Modgil, Iyad Rahwan, Chris Reed, Guillermo Simari, Matthew South, Gerard Vreeswijk, Steven Willmott, et al. Towards an Argument Interchange Format. *The Knowledge Engineering Review*, 21(04):293–316, 2006.
- [3] Rory Duthie, John Lawrence, Katarzyna Budzynska, and Chris Reed. The CASS Technique for Evaluating the Performance of Argument Mining. *to appear in Proceedings of the Third Workshop on Argument Mining. Association for Computational Linguistics*, 2016.
- [4] Alejandro J García and Guillermo R Simari. Defeasible logic programming: An argumentative approach. *Theory and practice of logic programming*, 4(1+ 2):95–138, 2004.
- [5] Thomas F Gordon, Henry Prakken, and Douglas Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10):875–896, 2007.
- [6] Paul A Kirschner, Simon J Buckingham-Shum, and Chad S Carr. *Visualizing argumentation: Software tools for collaborative and educational sense-making*. Springer, 2003.
- [7] John Lawrence, Floris Bex, Chris Reed, and Mark Snaith. AIFdb: Infrastructure for the Argument Web. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pages 515–516, 2012.
- [8] Marie-Francine Moens. Argumentation mining: Where are we now, where do we want to be and how do we get there? In *FIRE '13 Proceedings of the 5th 2013 Forum on Information Retrieval Evaluation*, 2013.
- [9] Andreas Peldszus and Manfred Stede. From argument diagrams to argumentation mining in texts: a survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1):1–31, 2013.
- [10] Brian Plüss and Anna De Liddo. Engaging citizens with televised election debates through online interactive replays. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, pages 179–184. ACM, 2015.
- [11] Mark Snaith, Rolando Medellin, John Lawrence, and Chris Reed. Arguers and the Argument Web. In *Proceedings of the 13th workshop on Computational Models of Natural Argument (CMNA 13)*, Rome, Italy, 2013. Springer.
- [12] Mark Snaith and Chris Reed. TOAST: Online ASPIC+ implementation. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA 2012)*, pages 390–397, Vienna, 2012. IOS Press.
- [13] Tim van Gelder. The rationale for Rationale. *Law, probability and risk*, 6(1-4):23–42, 2007.
- [14] Nina Wacholder, Smaranda Muresan, Debanjan Ghosh, and Mark Aakhus. Annotating multiparty discourse: Challenges for agreement metrics. *LAW VIII*, page 120, 2014.