Rationalisation of Profiles of Abstract Argumentation Frameworks[1]

Stéphane Airiau[a] Elise Bonzon[b] Ulle Endriss[c] Nicolas Maudet[d] Julien Rossit[b]

[a] *LAMSADE, Univiversité Paris-Dauphine, France*
[b] *LIPADE, Univiversité Paris Descartes, France*
[c] *ILLC, University of Amsterdam, The Netherlands*
[d] *LIP6, Université Pierre et Marie Curie, France*

Abstract

Different agents may have different points of view. This can be modelled using different abstract argumentation frameworks, each consisting of a set of arguments and a binary attack-relation between them. A question arising in this context is whether the diversity of views observed in such a profile of argumentation frameworks is consistent with the assumption that every individual framework is induced by a combination of, first, some basic factual attack-relation between the arguments and, second, the personal preferences of the agent concerned. We treat this question of *rationalisability* of a profile as an algorithmic problem and identify tractable and intractable cases. This is useful for understanding what types of profiles can reasonably be expected to come up in a multiagent system.

1 Introduction

The model of abstract argumentation introduced by Dung [3] is at the root of a vast amount of work in artificial intelligence and multiagent systems. In a nutshell, this model abstracts away from the content of an argument, and thus sees argumentation frameworks as directed graphs, where the nodes are arguments and the edges are attacks between arguments—in the sense that one argument undercuts or contradicts another argument. Different semantics provide principled approaches to selecting sets of arguments that can be viewed as coherent when taken together.

Starting with the work of Coste-Marquis et al. [2], in recent years, a number of authors have addressed the problem of aggregating several argumentation frameworks, each associated with the stance taken by a different individual agent, into a single collective argumentation framework that would appropriately represent the views of the group as a whole. This is an interesting and fruitful line of research, bringing together concerns in abstract argumentation with the methodology of social choice theory, but it raises one important question: For a given profile of argumentation frameworks, is it in fact conceivable that that profile would manifest itself? Intuitively speaking, it may often seem more natural to encounter a profile with similar individual attack-relations. So, how do we explain the differences in perspective of the individual agents for a given profile?

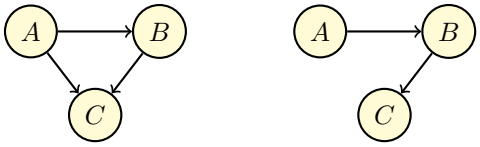
The point that the attack-relation should not be viewed as absolute and objective, but may very well depend on the individual circumstances of the agent considering the arguments in question, is central to the study of argumentation. Frameworks for modelling this phenomenon have been proposed by several authors. Here we adopt a preference-based approach, in the *value-based* variant originally due to Bench-Capon [1]. In his model, whether argument A ultimately defeats argument B does not only depend on whether A attacks B in an objective sense, but also on how we rank the importance of the social or moral values attached to A and B: If we rank the value associated with B strictly above that associated with A, we may choose to ignore any attacks of A on B.

[1] This is an extended abstract of a paper that appears in the *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2016)*.

At the technical level, we thus ask the following question: Given a profile of argumentation frameworks $(AF_1, \ldots, AF_n)$, one for each agent, can this profile be explained in terms of a single master argumentation framework, an association of arguments with values, and a profile of preference orders over values $(\succcurlyeq_1, \ldots, \succcurlyeq_n)$, one for each agent? In other words: Can the profile of argumentation frameworks observed be *rationalised*? To be able to answer this question in the affirmative, for every agent i, we require AF_i to be exactly the argumentation framework we obtain when the master argumentation framework with its associated values is reduced using the preference order $\succcurlyeq_i$. We may wish to impose any number of *constraints* on the solutions for this *rationalisability problem* we are interested in: e.g., constraints on the attack relation of the master argumentation framework, constraints on the number of values used for rationalisation, or constraints on the preference relations admitted.

2 Example

Suppose we observe two agents with the following argumentation frameworks:



Thus, they disagree on whether argument A attacks argument C (agent 1 says it does and agent 2 says it does not). As preferences can only cancel attacks rather than creating new ones, our best chance at rationalisation is to assume that the argumentation framework of agent 1 is also the master argumentation framework. Whatever values we end up associating with the arguments, if we assume that agent 1's preference relation is such that she is indifferent between all values, then we won't have to cancel any of the attacks for her, and rationalisation works correctly as far as she is concerned.

Hence, our original rationalisability problem for this example now reduces to the question of whether we can rationalise the righthand argumentation framework under the constraint that the lefthand argumentation framework is the master framework. If we associate each argument with a distinct value and if we permit incomplete preferences, then rationalisation is possible: If agent 2 prefers the value of C to the value of A and does not express a preference between any other pair, then the attack between A and C gets cancelled and all other attacks remain in place. Can we also rationalise using only *two* distinct values? No! We need two distinct values for A and C to cancel the attack between them. If B gets the same value as A, then we must also cancel the attack between B and C. And if B gets the same value as C, then we must also cancel the attack between A and B. Finally, can we also rationalise using a *complete* preference order? No! As we have seen before, agent 2 must strictly prefer the value of C to the value of A. Now, if we place the value of B anywhere above the value of A in the preference order, then we must also cancel the attack between A and B. On the other hand, if we place the value of B anywhere below the value of C, then we must also cancel the attack between B and C.

3 Results

Besides the introduction of the rationalisability problem itself, our contribution in the full paper consists in the development of algorithms to efficiently solve the rationalisability problem for a range of different constraints, and, for one choice of constraints, a complexity result showing that for those constraints the problem it is intractable. We also discuss possible applications in some detail.

References

[1] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[2] S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis. On the merging of Dung's argumentation systems. *Artificial Intelligence*, 171(10–15):730–753, 2007.

[3] 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):321–358, 1995.