

Dealing with the dynamics of proof-standard in argumentation-based decision aiding

Wassila Ouerdane and Nicolas Maudet and Alexis Tsoukias¹

Abstract. Usually, in argumentation, the proof-standards that are used are fixed a priori by the procedure. However (multicriteria) decision-aiding is a context where it may be modified dynamically during the process, depending on the responses of the decision-maker. The expert indeed needs to adapt and refine its choice of an appropriate method of aggregating arguments pros and cons, so that it fits the preference model inferred from the interaction. In this short paper we introduce how this aspect can be handled in an argumentation-based decision-aiding framework. The first contribution of the paper is conceptual: the notion of a concept lattice based on simple properties and allowing to navigate among the different proof-standards is put forward. We then show how this can be integrated within the Carneades model.

1 Introduction

From the seminal work of [4] to recent accounts [1], argumentation has been advocated as a relevant approach to account for decision-aiding. Under such a perspective, an agent faced with several possible candidate actions (or alternatives) will seek a decision which is “sufficiently” justified in the light of the arguments which can be constructed for or against the considered alternative. One (important) aspect of argumentation is to define what valid justifications (sets of arguments) can be considered, for instance whether reinstatement should be used or not [3, 9]. On top of that, though, a proper definition of what “sufficiently” means should be given, for in most situations conflicting justifications may typically be built for and against the considered decision. Doing that amounts to defining a *proof standard* [7]. In most applications (for example those dealing with legal issues), the proof standard to be used is fixed a priori, given exogenously by the procedure. In this paper, we are interested in (multicriteria) decision-aiding [2]: an expert helps a client in the process of choosing a single decision to be undertaken, given that the different actions may be evaluated on several dimensions (criteria). Part of the job of the expert is precisely to choose a given aggregation method [8] and adapt it to the responses provided by the client during the course of interaction, as illustrated on the following example. A decision maker (U) specifies his evaluation model, comprising of four actions $A = \{a_0, a_1, a_2, a_3\}$, five criteria $H = \{h_0, h_1, h_2, h_3, h_4\}$, and the following performance table:

	h_0	h_1	h_2	h_3	h_4
a_0	7	6	2	3	5
a_1	6	4	8	4	7
a_2	3	2	5	5	3
a_3	7	7	2	0	2

- (1) E: I recommend a_1 as being the best choice.
- (2) U: Why?
- (3) E: Because a_1 is preferred to each other alternatives by a majority of criteria: it is ranked first on h_2 and h_4 , and is only beaten by a_2 on h_3 . But a_1 beats a_2 on h_0 .
- (4) U: I see, but I would prefer a_0 to a_1
- (5) E: Why?
- (6) U: Because a_0 is better on $\{h_0, h_1\}$
- (7) E: Fine. But then why is not a_3 preferred?
- (8) U: No. a_3 is too bad on h_3 . This would not be justifiable.

The expert (E) makes a justified recommendation at turn (3). On turn (4) the decision maker puts forward the alternative that he feels intuitively as being the best here. The expert challenges this claim and at turn (6) the decision maker gives an explanation based on the fact that a_0 is better on two criteria: h_0 and h_1 . The expert again challenges the claim by emphasizing that on this basis a_3 should then be preferred. Finally the decision maker gives a reason for not accepting a_3 : a very low score on criteria h_3 . In both turn (6) and (8) the responses of the decision maker would trigger a difficult task for a system, since the assumptions made at the beginning of the dialogue (perhaps based on some initial elicitation) turn out to be inadequate, and invalidate the method chosen to make the final recommendation.

The difficulty to choose a procedure is mainly due to the fact that, at the beginning of the process, it is virtually impossible to have all the preferential information of the decision maker available. Moreover, these preferences may change and evolve during the interaction, leading to corrections or new information being provided. Still, a system should be able to decide on the fly whether a procedure is appropriate or not, and which one should be selected.

2 Navigating Through Proof Standards

Dealing with the dynamics of the proof standards poses a number of challenges. The first one is to define a principled way to identify an adequate proof standard, given the current state of the interaction. The second one is to design a mechanism that will allow to automatically update the proof as a result of the client’s responses.

How to identify a proof standard? The key idea here is to identify a set of properties that will help us to discriminate each procedure. Such properties will correspond to some characteristics of the decision maker’s preferences, corresponding to a set of conditions supporting the use of a given proof. For instance, *ordinality* means that only the ordering among actions is relevant for comparison, in particular the specific difference of performance values is not.

¹ LAMSADE, University Paris-Dauphine, France, email: {wassila.ouerdane,nicolas.maudet,alexis.tsoukias}@dauphine.fr

How to choose a proof standard? The idea here is just to favour methods that are *simple* for the user to understand. This notion is difficult to grasp in theory. Our solution is to let the system start with a favourite proof to be used: this can be provided by the expert, or sometimes by the decision maker himself (who may be familiar with a specific proof standard). This would typically be a *weighted sum* or a *simple majority*. The properties corresponding to this procedure are assumed to be true by default. Now as the user provides more preferential information (during the interaction), the system should be able to adapt and jump to the new favoured method. The navigation among the different candidate proof standards depends on the properties that are currently satisfied or contradicted. To account for that, we formalise the relationship between the set of properties and the set of proofs by a *Concept Lattice* as proposed in *Formal Concept Analysis* (FCA) [5]. In our context, a *formal concept* corresponds to a pair (R_i, P_i) , such that $R_i \subseteq R$ the set of proof standards and $P_i \subseteq P$ the set of properties. The navigation works as follows: position yourself in the lattice at the initially chosen proof (attached to this proof the property vector with values assumed to be as required by the corresponding proof). When some properties are contradicted by some decision maker's responses, the procedure moves to the appropriate node in the lattice. This gives a natural interpretation of how simple is a proof standard, namely how close (in terms of the number of properties that differ) it stands from the initially chosen proof standard.

Now we need more a detailed representation of the current state of the interaction, making explicit how the properties can be modified. The Carneades model [6] provides an adequate framework for this.

3 Integration in the Carneades Model

An extended acceptability function. The acceptability of a statement depends essentially on three elements: its dialectical status during the dialogue (*stated*, *questioned*, *accepted* and *rejected*), its proof standard and its *premises types*. Indeed, depending on their type, different requirement are attached to the premises considered. *Ordinary premises* must be supported with further ground, *assumptions* can be assumed to hold until they are questioned, and finally *exceptions* don't hold in the absence of evidence to the contrary. Determining the acceptability of a statement is eventually the result of a procedure which recursively determine the acceptability of its premises on the basis of the elements mentioned above [6]. In our case, an argument graph is constructed with the arguments exchanged during the decision aiding process, including an explicit representation of the proof standard, so that it can be (indirectly) discussed and challenged as any other statement in the graph. Technically, a new link is added to represent the fact that a given proof is assigned to a statement. The proof standard itself is a statement of a special type, to which are attached different properties, as discussed in the previous section.

The acceptability function of [6] is extended in an obvious way: now the system requires that the statement satisfies the proof standard attached to it, and also that the proof standard is itself acceptable.

Infinite recursion is avoided because for the statement nodes of type "proof standard", a unique way to assess the acceptability is imposed. A single argument will support the use of a proof standard. The properties are conceived as a set of premises of this argument supporting "theoretically" the use of a specific proof standard, and contradicting one of this property will invalidate the acceptability of the proof standard.

Critical responses, positive and negative evidence. To each of these premises is now attached a list of possible *critical responses*, which play the same role as critical questions (a set of questions that represent attacks, challenges or criticisms to a given argument scheme). Critical responses play the same role, except that the decision maker is typically not aware that it constitutes a counter-argument to some claim. These responses correspond to typical manners that the decision maker may contradict a given property. There is an important difference to make between two types of responses:

- *Positive Evidence (PE)*—provides supporting evidence to the fact that the property is satisfied. This would typically consist of general statements that the decision maker may make, e.g. "the values in the table are not relevant to me" would be positive evidence for ordinality.
- *Negative Evidence (NE)*—provides an explicit counter-example to a specific property, e.g. "frankly the score of this alternative x is so much better than the one of y that I can't take y " would be negative evidence for ordinality.

Following [6], critical responses will be modeled as premises. The type of premises used here depends on the kind of evidence required. As explained earlier, negative evidence is sufficiently convincing since it provides an explicit counter-example. They will be modeled as *assumptions*. On the other hand, positive evidence may require further ground to be acceptable (the decision maker is *not* supposed to be competent to state general features of his preference). They will be modeled as *ordinary premises*. The system may subsequently test the decision-maker with a series of question to establish some grounds upon which the validity of the claim may be granted via a *property testing subdialogue*. The idea of a property testing subdialogue is that the system enters into an embedded dialogue with the objective to back a positive evidence claim. Again, as the decision maker himself is not in a position to provide some reasons to back this claim, the process will be guided by the system.

Integrating the procedure. The different pieces mentioned so far are glued together into a general procedure which makes explicit how our proposal can be integrated into a Carneades-based dialogue game. In particular, it ensures that the status of the different properties can only be modified a finite number of times.

REFERENCES

- [1] L. Amgoud and H. Prade, 'Using arguments for making and explaining decisions', *Journal of Artificial Intelligence*, **173**, 413–436, (2009).
- [2] D. Bouyssou, Th. Marchant, M. Pirlot, A. Tsoukiàs, and Ph. Vincke, *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*, 2006.
- [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).
- [4] J. Fox and S. Parsons, 'On Using Arguments for Reasoning about Actions and Values', in *Proc. of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*, pp. 55–63, (1997).
- [5] B. Ganter and R. Willer, *Formal concept Analysis*, Springer, 1999.
- [6] T. Gordon, H. Prakken, and D. Walton, 'The Carneades model of argument and burden of proof', *Artificial Intelligence*, **171**(4), 875–896, (2007).
- [7] T.F. Gordon and D. Walton, 'Proof burdens and standards', in *Argumentation in Artificial Intelligence*, eds., I. Rahwan and G. Simari, (2009).
- [8] A. Guitouni and J.M. Martel, 'Tentative guidelines to help choosing an appropriate MCDA method', *European Journal of Operational Research*, **109**(2), 501–521, (1998).
- [9] J. F. Horty, 'Argument construction and reinstatement in logics for defeasible reasoning', *Artificial Intelligence and Law*, **9**(1), 1–28, (2001).