On the role of preferences in argumentation frameworks

Leila Amgoud
IRIT - CNRS
118, route de Narbonne,
31062 Toulouse Cedex 9 France
amgoud@irit.fr

Srdjan Vesic
IRIT - Université de Toulouse
118, route de Narbonne,
31062 Toulouse Cedex 9 France
vesic@irit.fr

***Abstract*—The aim of this paper is to study how *preferences*, which are used to model intrinsic strengths of arguments, can be used in argumentation. We show that they play two roles: i) to repair the attack relation between arguments, and ii) to refine the evaluation of arguments. Then, we point out that the existing approaches for preference-based argumentation model only the first role. They may also return non conflict-free extensions. We propose a general framework that overcomes those limitations.**

I. Introduction

Argumentation is a reasoning model based on the construction and the evaluation of interacting *arguments*. One of the most popular argumentation frameworks (AF) has been proposed by Dung in [6]. It consists of a set of arguments and an attack relation among them. The attack relation is at the heart of all Dung's semantics. In [1], [3], it has been argued that other criteria can be taken into account for evaluating arguments. Namely, the strengths of arguments may play a key role in the evaluation process. The idea is that an attack may fail if the attacker is weaker than the attacked argument. Such attacks will be referred to as *critical* attacks. Extensions of Dung's framework by preferences have been proposed in the literature. The basic idea behind them is to remove *critical* attacks from the graph of attacks and to apply Dung's semantics on the remaining sub-graph. More recently (in [5]) the authors defined a new semantics which uses preferences in a different manner, namely to refine the result of an argumentation framework.

In this paper, we study the difference between the roles played by preferences in [1], [3] and in [5] and for the first time we propose a general system that models both roles of preferences. We argue that preferences between arguments play two roles: i) to weaken attacks, and ii) to refine the evaluation of arguments. These roles are somehow independent and both should be modeled in order to have correct and refined results.

Regarding the first role, we propose a novel approach that overcomes the limitations of the one followed in [1], [3]. The idea is to invert the arrows of critical attacks instead of removing them. In fact, by removing a critical attack we lose an important information which is the fact that the two arguments of this attack should not be accepted simultaneously. This leads in some cases to conflicting extensions of arguments. Inverting arrows allows us to encode the conflict between the two arguments and in the same time the preference between them. We show that our approach is well-founded in the sense that it guarantees safe and intuitive results. Regarding the second role, we show that the refinement is done via a preference relation defined on the power set of the set of arguments. This relation is not unique. However, it should satisfy some basic properties (like reflexivity and transitivity). Finally, we define an abstract *rich preference-based argumentation framework* in which the two roles of preferences are captured.

The paper is organized as follows: Section II recalls Dung's AF. Section III discusses the two roles of preferences and how they should be modeled. Section IV presents an abstract framework in which the two roles are formalized. Section V compares our approach with existing works. The last section concludes.

II. Basics of argumentation

The argumentation framework is defined as follows:

Definition 1 (Argumentation framework [6]): An *argumentation framework* (AF) is a pair $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, where $\mathcal{A}$ is a set of arguments and $\mathcal{R}$ is an attack relation ($\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$). The notation $a\mathcal{R}b$ means that a *attacks* b.

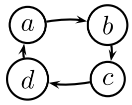
Definition 2 (Conflict-free, Defence [6]): Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{B} \subseteq \mathcal{A}$.

- $\mathcal{B}$ is *conflict-free* iff $\nexists a, b \in \mathcal{B}$ s.t. $a\mathcal{R}b$.
- $\mathcal{B}$ *defends* an argument a iff $\forall b \in \mathcal{A}$ s.t. $b\mathcal{R}a$, $\exists c \in \mathcal{B}$ s.t. $c\mathcal{R}b$.

Definition 3 (Acceptability semantics [6]): Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{B} \subseteq \mathcal{A}$ is conflict-free.

- $\mathcal{B}$ is an *admissible* extension iff it defends all its elements.
- $\mathcal{B}$ is a *complete* extension iff it defends its elements and contains all the arguments that it defends.
- $\mathcal{B}$ is a *grounded* extension iff it is the minimal (for set inclusion) complete extension.
- $\mathcal{B}$ is a *preferred* extension iff it is a maximal (for set inclusion) admissible extension.
- $\mathcal{B}$ is a *stable* extension iff it is a preferred extension that attacks any element in $\mathcal{A} \setminus \mathcal{B}$.

Example 1: Let us consider the AF depicted below.



This AF has two preferred and stable extensions: $\{a, c\}$ and $\{b, d\}$. Its grounded extension is the empty set.

Definition 4: Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\mathcal{E}_1, \ldots, \mathcal{E}_n$ its extensions (under a given semantics). Let $a \in \mathcal{A}$.

- a is *skeptically accepted* iff $\forall i = 1, \ldots, n$, $a \in \mathcal{E}_i$.
- a is *credulously accepted* iff $\exists i = 1, \ldots, n$ s.t. $a \in \mathcal{E}_i$.
- a is *rejected* iff $\forall i = 1, \ldots, n$, $a \notin \mathcal{E}_i$.

Example 2 (Example 1 Cont.): The arguments a, b, c, d are credulously accepted under stable and preferred semantics, while they are all rejected under grounded semantics.

III. The role of preferences in argumentation

In what follows, we assume that $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ is an arbitrary argumentation framework where $\mathcal{A}$ is *finite*. Let $\geq$ be a binary relation that expresses preferences between arguments of $\mathcal{A}$. For instance, an argument may be preferred to another if it is grounded on more certain information, or if it promotes a more important value. Throughout the paper, the relation $\geq \subseteq \mathcal{A} \times \mathcal{A}$ is assumed to be a preorder[1]. For two arguments a and b, writing $a \geq b$ (or $(a, b) \in \geq$) means that a is at least as strong as b. We write $a > b$ iff $a \geq b$ and not ($b \geq a$).

We distinguish two roles of preferences:

1) To identify and handle the critical attacks in an AF.
2) To refine the evaluation of arguments.

Next subsections discuss in detail each of these roles, their links and how they can be modeled. Finally, we propose an abstract model that extends Dung's AF by preferences between arguments. The model integrates both roles of preferences.

A. Handling critical attacks

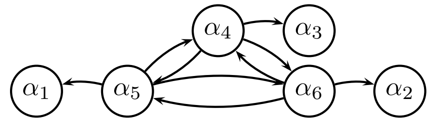
An attack from an argument b towards an argument a always wins unless b is itself attacked by another argument. However, this assumption is very strong because some attacks may be *critical* and cannot always "survive".

Definition 5 (Critical attack): Let $\mathcal{F} = (\mathcal{A}, \mathcal{R})$ be an AF and $\geq \subseteq \mathcal{A} \times \mathcal{A}$. An attack $(b, a) \in \mathcal{R}$ is *critical* iff $a > b$.

Example 3: Let $\Sigma = \{a, \neg b, a \rightarrow b\}$ be a propositional knowledge base s.t. a is *more certain* than the two other formulas. The following arguments[2] are built from this base:

$\alpha_1 : (\{a\}, a)$	$\alpha_2 : (\{\neg b\}, \neg b)$
$\alpha_3 : (\{a \rightarrow b\}, a \rightarrow b)$	$\alpha_4 : (\{a, \neg b\}, a \wedge \neg b)$
$\alpha_5 : (\{\neg b, a \rightarrow b\}, \neg a)$	$\alpha_6 : (\{a, a \rightarrow b\}, b)$

Below are depicted the attacks wrt assumption attack[3] [7].



The above framework has three stable extensions: $\mathcal{E}_1 = \{\alpha_1, \alpha_2, \alpha_4\}$, $\mathcal{E}_2 = \{\alpha_2, \alpha_3, \alpha_5\}$, and $\mathcal{E}_3 = \{\alpha_1, \alpha_3, \alpha_6\}$. None of the six arguments is skeptically accepted. Arguments α_1 and α_5 are both credulously accepted. Thus, both formulas a and $\neg a$ can be inferred from Σ. The result is counter-intuitive since a is more certain than $\neg a$. Thus, we would expect to infer the formula a.

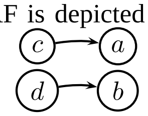
This example shows that preferences between arguments should be taken into account. For instance, α_1 is stronger than α_5. Consequently, the attack from α_5 to α_1 should fail. This means that the relation $\geq$ should take precedence over the attack relation $\mathcal{R}$ in case of critical attacks.

Note that preferences do not always take precedence over attacks, as illustrated by the following example.

Example 4: Assume the following witnesses:

- w_1: I know that 2 men entered the building at 11:00 am since I was in front of the building waiting for Miss Jones from 10:50 am to 11:15 am (**a**).
- w_2: I saw witness w_1 in the restaurant at 11:05 am (**c**).
- w_3: Mr Smith could not see Miss Jones that morning in her office since he was in California (**b**).
- w_4: Mr Smith was in New York that morning (**d**).

The corresponding AF is depicted in the figure below.



The set $\{c, d\}$ is the only grounded, preferred and stable extension. Assume now that the witness w_1 is more reliable than w_4 and that w_3 is more trustworthy than w_2. Thus, we have $a > d$ and $b > c$. According to these preferences, the set $\{a, b\}$ is better than the extension $\{c, d\}$ since the former contains the arguments advanced by the most reliable witnesses. However, we should accept $\{c, d\}$ and reject $\{a, b\}$. The fact that a is stronger than d does not help a to be defended against c. The two arguments c and d are on completely different topics. Thus, preferences do not take precedence over attacks in this case.

Conclusion 1: In an AF, the preference relation *should* take precedence over the attack relation when handling critical attacks while the attack relation should be privileged when handling safe attacks.

This role of preferences (i.e. handling critical attacks) has already been identified in the literature, namely in [1], [3]. The basic idea behind those works is to remove the critical attacks from the argumentation graph and to apply Dung's semantics on the remaining sub-graph. Unfortunately, this approach suffers from a great drawback when the attack relation is not symmetric.

Example 5 (Example 3 Cont.): Assume that arguments are compared using the *weakest link principle*[4] [4]. According

[1] A binary relation is a *preorder* iff it is *reflexive* and *transitive*.

[2] An *argument* is a pair (H, h) where H is its *support* and h its *conclusion*. H is a minimal subset of Σ that is consistent and infers classically h.

[3] An argument α *attacks* β iff the conclusion of α is the contrary of a formula in the support of β.

[4] An argument α is *preferred* to β if the least certain formula in the support of α is more certain than the least certain formula in the support of β.

to this relation, the argument α_1 is strictly preferred to the others, which are themselves equally preferred. The classical approaches of PAFs remove the attack from α_5 to α_1 and get $\{\alpha_1, \alpha_2, \alpha_3, \alpha_5\}$ as a stable extension. Note that this extension, which intends to support a *coherent point of view*, is conflicting since it supports both a and $\neg a$[5].

The main reason behind this dysfunction is that by removing an attack, an important information – conflict between the arguments – is lost. Consequently, existing approaches do not guarantee that extensions are conflict-free. In what follows, we propose a novel approach whose main idea is to modify the graph of attacks in such a way that, for any critical attack, the preference between the arguments is taken into account and the conflict between the two arguments of the attack is represented. For this purpose, we *invert* the arrow of the critical attack.

Definition 6 (PAF): A *preference-based argumentation framework* (PAF) is a tuple $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation and $\geq$ is a partial preorder on $\mathcal{A}$. The extensions of $\mathcal{T}$ under a given semantics are the extensions of the argumentation framework $(\mathcal{A}, \mathcal{R}_r)$, called *repaired framework*, under the same semantics with: $\mathcal{R}_r = \{(a, b) | (a, b) \in \mathcal{R}$ and not $(b > a)\} \cup \{(b, a) | (a, b) \in \mathcal{R}$ and $b > a\}$.

Example 6 (Example 5 Cont.): The repaired framework, in which the arrow from α_5 to α_1 is replaced by another arrow emanating from α_1 towards α_5, has two stable extensions: $\mathcal{E}'_1 = \{\alpha_1, \alpha_2, \alpha_4\}$, and $\mathcal{E}'_2 = \{\alpha_1, \alpha_3, \alpha_6\}$. Note that both extensions support coherent points of view. Moreover, α_1 is skeptically accepted, thus a should be inferred from Σ.

From Definition 6, it is clear that if a PAF has no critical attacks, then the repaired framework coincides with the basic one. This also shows that when a PAF has no critical attacks, then preferences do not play any role in the evaluation process. Our approach does not suffer from the drawback of the existing one, since it always delivers conflict-free extensions of arguments.

Another important result is that the fact of inverting the arrows of critical attacks in an argumentation graph does not affect the status of arguments that are not related to the arguments of those attacks. This means that our approach has no bad side effects.

Our approach privileges the strongest arguments of a PAF. Before presenting the formal result, let us define the strongest arguments.

Definition 7 (Maximal elements): Let $\mathcal{O}$ be a set of objects and $\geq \subseteq \mathcal{O} \times \mathcal{O}$ is a (partial or total) preorder. The *maximal elements* of $\mathcal{O}$ wrt. $\geq$ are $\mathtt{Max}(\mathcal{O}, \geq) = \{o \in \mathcal{O} \mid \nexists o' \in \mathcal{O}$ s.t. $o' > o\}$.

Property 1: Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t $\geq$ is complete[6]. If $\mathtt{Max}(\mathcal{A}, \geq)$ is conflict-free (wrt. $\mathcal{R}$), then $\forall a \in \mathtt{Max}(\mathcal{A}, \geq)$, a is skeptically accepted in $\mathcal{T}$ wrt. preferred and grounded semantics. If $\mathcal{T}$ has at least one stable extension, then a is skeptically accepted wrt. stable semantics.

The following result shows that when the preference relation $\geq$ is a linear order (i.e. reflexive, antisymmetric, transitive and complete), then the corresponding PAF has a unique stable/preferred extension.

Proposition 1: Let $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq)$ be a PAF s.t. $\mathcal{R}$ is irreflexive, $\geq$ is a linear order and let $n = |\mathcal{A}|$.

- $\mathcal{T}$ has exactly one stable extension.
- Stable, preferred and grounded extensions of $\mathcal{T}$ coincide.
- This extension is computed in $\mathcal{O}(n^2)$ time.

In the case when the attack relation is symmetric, our approach returns the same result as the approach developed in [1], [3]. We can also show that when the attack relation is symmetric, the extensions of a PAF are a subset of those of its basic framework. This means that preferences filter the extensions. This does not hold in case the attack relation is not symmetric.

When the attack relation is symmetric and irreflexive, the corresponding PAF is *coherent* (i.e. its preferred and stable extensions coincide) and it has at least one stable extension.

B. *Refining AFs by preferences*

The argumentation framework of Example 4 has no critical attacks. It seems that the two preferences $a > d$ and $b > c$ are useless in this case. One wonders whether this is always the case when there are no critical attacks in an AF. The following example shows that there are situations in which preferences are useful, namely for refining the results returned by the acceptability semantics. This is the second role that preferences may play in an AF.

Example 7 (Example 2 Cont.): Let us assume that $a > b$ and $c > d$. Note that any element of $\{b, d\}$ is weaker than at least one element of $\{a, c\}$. Thus, it is natural to consider $\{a, c\}$ as better than $\{b, d\}$. Consequently, we may conclude that the two arguments a and c are "more acceptable" than b and d. This is important, in particular in a decision making problem. Assume that the two arguments c, d support an option o_1 while the two arguments b, d support another option, say o_2. Since, only one option will be chosen at the end, the preferences make it possible to select o_1.

Note that a refinement amounts to *compare* subsets of $\mathcal{A}$. In Example 1, the *democratic* relation, $\succeq_d$, was used: for $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{A}$, $\mathcal{E} \succeq_d \mathcal{E}'$ iff $\forall x' \in \mathcal{E}' \setminus \mathcal{E}, \exists x \in \mathcal{E} \setminus \mathcal{E}'$ s.t. $x > x'$. The relation $\succeq_d$ is not the only possible comparison relation:

Example 8: Let $\mathcal{A} = \{a, b, c\}$, $\mathcal{R} = \{(a, b), (a, c), (c, a)\}$ and $\geq = \{(a, b)\}$. This system has two extensions under preferred or stable semantics: $\mathcal{E}_1 = \{a\}$ and $\mathcal{E}_2 = \{b, c\}$. According to $\succeq_d$, $\mathcal{E}_1$ and $\mathcal{E}_2$ are not comparable. However, it can be argued that $\mathcal{E}_1$ is better than $\mathcal{E}_2$ since for every argument in $\mathcal{E}_1 \setminus \mathcal{E}_2$, there exists a strictly weaker argument in $\mathcal{E}_2 \setminus \mathcal{E}_1$. In the previous example, the *elitist* relation was used for comparing the two extensions: for $\mathcal{E}, \mathcal{E}' \subseteq \mathcal{A}$, $\mathcal{E} \succeq_e \mathcal{E}'$ iff $\forall x \in \mathcal{E} \setminus \mathcal{E}', \exists x' \in \mathcal{E}' \setminus \mathcal{E}$ s.t. $x > x'$.

[5] This problem is not due to "forgotten" or "missing" arguments. No matter how many arguments (constructed from Σ) are added, there will always be an extension containing α_1 and α_5.

[6] A relation $\geq$ on a set $\mathcal{A}$ is complete iff for all $a, b \in \mathcal{A}$, $a \geq b$ or $b \geq a$.

Definition 8 (Refinement relation): Let $(\mathcal{A}, \geq)$ be s.t. $\mathcal{A}$ is a set of arguments and $\geq \subseteq \mathcal{A} \times \mathcal{A}$ is a preorder. A *refinement relation,* denoted by $\succeq$, is a binary relation on $\mathcal{P}(\mathcal{A})^7$ s.t.

- $\succeq$ is reflexive and transitive
- For all $\mathcal{E} \subseteq \mathcal{A}$, if $a > b$, then $\mathcal{E} \cup \{a\} \succ \mathcal{E} \cup \{b\}$

Property 2: Relations $\succeq_d$ and $\succeq_e$ are refinement relations.

IV. Rich PAFs

In the previous section, we have shown that preferences may play two roles in an AF: handling critical attacks and refining the results of an AF. The question now is what are the links between the two roles? Is it possible that one of them recovers the other?

We see from Example 1 that refining may be necessary and cannot be omitted. Is the refinement alone enough to obtain the desirable result? The following example provides a negative answer on that question.

Example 9: Let us consider the argumentation framework depicted in the left side of the figure below.

Assume that $a > b$ and $a > d$ and $d > c$. This framework has two stable/preferred extensions: $\{a, c\}$ and $\{b, d\}$. If we apply the democratic relation $\succeq_d$, the set $\{a, c\}$ is clearly preferred to $\{b, d\}$. Thus, the two arguments a and c would be skeptically accepted. However, this result is not intuitive. The reason is that d defends itself against its unique attacker c. Thus, d should be accepted and consequently, c should be rejected, thus the expected extension would be: $\{a, d\}$. Note that $\{b, d\}$ cannot be an extension since b is attacked by a stronger argument (a). By inverting the arrows of the critical attacks, we get the framework depicted in the right side of the above figure. This framework has a unique stable extension which is the expected result $\{a, d\}$.

Conclusion 2: Taking into account the preferences in the evaluation of arguments should be a two-step process:

1) Repairing the attack relation $\mathcal{R}$ by computing $\mathcal{R}_r$.
2) Refining the results of the framework $(\mathcal{A}, \mathcal{R}_r)$ by comparing its extensions using a refinement relation.

In what follows, we propose an abstract framework, called *rich preference-based argumentation framework*, in which both roles of preferences are modeled.

Definition 9 (Rich PAFs): A *rich PAF* is a tuple $\mathcal{T} = (\mathcal{A}, \mathcal{R}, \geq, \succeq)$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation, $\geq \subseteq \mathcal{A} \times \mathcal{A}$ is a preorder and $\succeq \subseteq \mathcal{P}(\mathcal{A}) \times \mathcal{P}(\mathcal{A})$ is a refinement relation. The extensions of $\mathcal{T}$ under a given semantics are the elements of $\mathtt{Max}(\mathcal{S}, \succeq)$ where $\mathcal{S}$ is the set of extensions (under the same semantics) of the PAF $(\mathcal{A}, \mathcal{R}, \geq)$.

Example 10 (Example 7 Cont.): Let us use the democratic relation $\succeq_d$. In this framework, $\mathcal{R}_r = \mathcal{R}$. The extensions of the rich PAF are $\mathtt{Max}(\{\{a, c\}, \{b, d\}\}, \succeq_d) = \{\{a, c\}\}$.

[7] $\mathcal{P}(\mathcal{A})$ denotes the powerset of a set $\mathcal{A}$.

V. Related work

Introducing preferences in argumentation frameworks goes back to Simari and Loui ([8]) where the authors have defined an AF in which arguments are built from a propositional knowledge base. Arguments grounded on specific information are stronger than the ones built from more general information. That idea has been generalized in [1] to any AF and to any preference relation. Unfortunately, the approach followed in [1] may return conflicting extensions when the attack relation is not symmetric. Our approach overcomes these limits. Furthermore, it is more general since it models even the second role of preference (i.e. the refinement).
In [2], only the first role of preferences is modeled, and only for three semantics, while the system proposed in this paper not only refines the result but can also be used with any semantics.
To the best of our knowledge, the only work on refinement is that appeared in [5]. The authors have proposed a particular refinement relation in case of stable semantics. In this sense, our work is more general since it accepts any refinement relation. Moreover, there is no restriction to particular semantics.

VI. Conclusion

The paper has presented a study on the role that preferences can play in an AF. Two roles are distinguished. The first one consists of identifying critical attacks. We have proposed a new approach for modeling this role and which overcomes the limitations of existing approaches. The basic idea is to invert the arrow of each critical attack instead of removing it. We have shown that such an approach is well-founded. The second role consists of refining the results of an AF. We have shown that a refinement amounts to compare, using a refinement relation, the extensions under a given semantics of an AF. We have thus proposed an abstract framework, called rich PAF, in which the two roles are modeled. The idea is to repair first the critical attacks, then to apply Dung's acceptability semantics on the repaired framework, and finally to apply a refinement relation on the extensions.

References

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Int.*, 34:197–216, 2002.
[2] L. Amgoud and S. Vesic. Repairing preference-based argumentation systems. In *IJCAI'09*, pages 665–670, 2009.
[3] T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *J. of Logic and Computation*, 13(3):429–448, 2003.
[4] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of UAI'93*, pages 411–419, 1993.
[5] Y. Dimopoulos, P. Moraitis, and L. Amgoud. Extending argumentation to make good decisions. In *ADT'09*, pages 225–236. LNCS 5783, 2009.
[6] P. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *AIJ*, 77:321–357, 1995.
[7] M. Elvang-Gøransson, J. Fox, and P. Krause. Acceptability of arguments as 'logical uncertainty'. In *Proceedings of ECSQARU'93*, number 747, pages 85–90. 1993.
[8] G. Simari and R. Loui. A mathematical treatment of defeasible reasoning and its implementation. *AIJ*, 53:125–157, 1992.