Preference Reasoning for Argumentation: Non-monotonicity and Algorithms

Souhila Kaci
CRIL
Rue de l'Université SP 16
62307 Lens France
kaci@cril.univ-artois.fr

Leendert van der Torre
ILIAS
University of Luxembourg
Luxembourg
leon.vandertorre@uni.lu

Abstract

In this paper we are interested in the role of preferences in argumentation theory. To promote a higher impact of preference reasoning in argumentation, we introduce a novel preference-based argumentation theory. Using non-monotonic preference reasoning we derive a Dung-style attack relation from a preference specification together with a defeat relation. In particular, our theory uses efficient algorithms computing acceptable arguments via a unique preference relation among arguments from a preference relation among sets of arguments.

Introduction

Dung's theory of abstract argumentation (Dung 1995) is based on a set of arguments and a binary attack relation defined over the arguments. Due to this abstract representation, it can and has been used in several ways, which may explain its popularity in artificial intelligence. It has been used as a general framework for non-monotonic reasoning, as a framework for argumentation, and as a component in agent communication, dialogue, decision making, *etc*. Dung's abstract theory has been used mainly in combination with more detailed notions of arguments and attack, for example arguments consisting of rules, arguments consisting of a justification and a conclusion, or attack relations distinguishing rebutting and undercutting. However, there have also been several attempts to modify or generalize Dung's theory, for example by introducing preferences (Amgoud & Cayrol 2002; Kaci, van der Torre, & Weydert 2006), defeasible priorities (Prakken & Sartor 1997; Poole 1985; Simari & Loui 1992; Stolzenburg *et al.* 2003), values (Bench-Capon 2003), or collective arguments (Bochman 2005).

In this paper we are interested in the role of preference reasoning in Dung's argumentation theory. An example from political debate has been discussed by Bench-Capon *et al.* (Atkinson, Bench-Capon, & McBurney 2005), where several arguments to invade Iraq are related to values such as respect for life, human rights, good world relations, and so on. In this paper we use a less controversial example to illustrate our theory where several arguments used in a debate between parents and their children are used to promote values like staying healthy, doing well at school, and so on.

In our theory, we integrate two existing approaches (though our approach differs both conceptually and technically from these approaches in several significant ways, as explained in the related work).

- We consider a preference based argumentation theory consisting of a set of arguments, an attack relation, and a preference relation over arguments. Then, like Amgoud and Cayrol (Amgoud & Cayrol 2002), we transform this preference based argumentation theory to Dung's theory, by stating that an argument A attacks another argument B in Dung's theory, when A attacks B in the preference-based theory, and B is not preferred to A. To distinguish the two notions of attack, we call the notion of attack in the preference-based theory *defeat*. The defeat and preference relation may be considered as an alternative representation of Dung's attack relation.
- Like Bench-Capon (Bench-Capon 2003), we consider value based argumentation, in which arguments are used to promote a value, and in which values are ordered by a preference relation. Moreover, in contrast to Bench-Capon, we use non-monotonic preference reasoning to reduce the ordered values to a preference relation over arguments. In analogy with the above, we say that the ordered values represent the preference relation over arguments.

Summarizing, starting with a set of arguments, a defeat relation, and an ordered set of values, we use the ordered values to compute a preference relation over arguments, and we combine this preference relation with the defeat relation to compute Dung's attack relation. Then we use any of Dung's semantics to define the acceptable set of arguments. In contrast to most other approaches (Amgoud & Cayrol 2002; Prakken & Sartor 1997; Poole 1985; Simari & Loui 1992; Stolzenburg *et al.* 2003) (but see (Amgoud, Parsons, & Perrussel 2000) for an exception), our approach to reason about preferences in argumentation does not refer to the internal structure of the arguments. We study the following research questions:

1. How to reason about ordered values and to derive a preference relation over arguments?
2. How to combine the two steps of our approach to directly define the acceptable set of arguments from a defeat relation and an ordered set of values?

To reason about ordered values and to compute the preference relation over arguments, we are inspired by insights from the non-monotonic logic of preference (Kaci & van der Torre 2005). When value v_1 is promoted by the arguments $A_1, \ldots, A_n$, and value v_2 is promoted by arguments $B_1, \ldots, B_m$, then the statement that value v_1 is preferred to value v_2 means that the set of arguments $A_1, \ldots, A_n$ is preferred to the set of arguments $B_1, \ldots, B_m$. In other words, the problem of reducing ordered values to a preference relation comes down to reducing a preference relation over sets of arguments to a preference relation over single arguments. We use both so-called optimistic and pessimistic reasoning to define the preference relation.

For the combined approach, we restrict ourselves to Dung's grounded semantics. For this semantics, we introduce an algorithm that shows how the computation of set of acceptable arguments can be combined with the optimistic reasoning to incrementally define the set of acceptable arguments, and we show why this works less well for pessimistic reasoning.

The layout of this paper is as follows. After presenting Dung's abstract theory of argumentation, and its extension to the preference-based argumentation framework, we introduce our value based argumentation theory, and show how to reduce a value based argumentation theory to a preference-based argumentation theory using optimistic or pessimistic reasoning. Then we introduce an algorithm for directly computing the set of acceptable arguments using grounded semantics. We also present an algorithm for ordering the arguments following the pessimistic reasoning. Lastly we discuss related work and conclude.

Abstract argumentation

Argumentation is a reasoning model based on constructing arguments, determining potential conflicts between arguments and determining acceptable arguments.

Dung's argumentation framework

Dung's framework (Dung 1995) is based on a binary attack relation among arguments.

Definition 1 (Argumentation framework) *An argumentation framework is a tuple $\langle \mathcal{A}, \mathcal{R} \rangle$ where $\mathcal{A}$ is a set of arguments and $\mathcal{R}$ is a binary attack relation defined on $\mathcal{A} \times \mathcal{A}$.*

We restrict ourselves to *finite* argumentation frameworks, i.e., when the set of arguments $\mathcal{A}$ is *finite*.

Definition 2 (Defence) *A set of arguments $\mathcal{S}$ defends A if for each argument B of $\mathcal{A}$ which attacks A, there is an argument C in $\mathcal{S}$ which attacks B.*

Definition 3 (Conflict-free) *Let $\mathcal{S} \subseteq \mathcal{A}$. The set $\mathcal{S}$ is conflict-free iff there are no $A, B \in \mathcal{S}$ such that $A\mathcal{R}B$.*

The following definition summarizes different acceptable semantics of arguments proposed in the literature:

Definition 4 (Acceptability semantics) *Let $\mathcal{S} \subseteq \mathcal{A}$.*

- *$\mathcal{S}$ is* admissible *iff it is conflict-free and defends all its elements.*
- *A conflict-free $\mathcal{S}$ is a* complete extension *iff $\mathcal{S} = \{A \mid \mathcal{S}$ defends $A\}$.*
- *$\mathcal{S}$ is a* grounded extension *iff it is the smallest (for set inclusion) complete extension.*
- *$\mathcal{S}$ is a* preferred extension *iff it is the largest (for set inclusion) complete extension.*
- *$\mathcal{S}$ is a* stable extension *iff it is a preferred extension that attacks all arguments in $\mathcal{A} \backslash \mathcal{S}$.*

The output of the argumentation framework is derived from the set of selected acceptable arguments w.r.t. an acceptability semantics.

Preference-based argumentation framework

An extended version of Dung's framework (Dung 1995) has been proposed in (Amgoud & Cayrol 2002) where a preference relation is defined on the set of arguments on the basis of the evaluation of arguments. We start with some definitions concerning preferences.

Definition 5 *A pre-order on a set $\mathcal{A}$, denoted $\succeq$, is a reflexive and transitive relation. $\succeq$ is total if it is complete and it is partial if it is not. The notation $A_1 \succeq A_2$ stands for A_1 is at least as preferred as A_2. $\succ$ denotes the order associated with $\succeq$. We write $\max(\succeq, \mathcal{A})$ for $\{B \in \mathcal{A}, \nexists B' \in \mathcal{A}$ s.t.$B' \succ B\}$ and we write $\min(\succeq, \mathcal{A})$ as $\{B \in \mathcal{A}, \nexists B' \in \mathcal{A}$ s.t. $B \succ B'\}$.*

Definition 6 illustrates how a total pre-order on $\mathcal{A}$ can also be represented by a well ordered partition of $\mathcal{A}$. This is an equivalent representation, in the sense that each total preorder corresponds to one ordered partition and vice versa. This equivalent representation as an ordered partition makes some definitions easier to read.

Definition 6 (Ordered partition) *A sequence of sets of arguments of the form $(E_1, \cdots, E_n)$ is the ordered partition of $\mathcal{A}$ w.r.t. $\succeq$ iff*

- *$E_1 \cup \cdots \cup E_n = \mathcal{A}$,*
- *$E_i \cap E_j = \emptyset$ for $i \neq j$,*
- *$\forall A, B \in \mathcal{A}$, $A \in E_i$ and $B \in E_j$ with $i < j$ iff $A \succ B$.*

An ordered partition of $\mathcal{A}$ is associated with pre-order $\succeq$ on $\mathcal{A}$ iff $\forall A, B \in \mathcal{A}$ with $A \in E_i, B \in E_j$ we have $i \leq j$ iff $A \succeq B$.

Definition 7 (Preference-based argumentation framework) *A preference-based argumentation framework is a triplet $\langle \mathcal{A}, \mathcal{D}, \succeq \rangle$ where $\mathcal{A}$ is a set of arguments, $\mathcal{D}$ is a binary defeat relation defined on $\mathcal{A} \times \mathcal{A}$ and $\succeq$ is a (total or partial) pre-order (preference relation) defined on $\mathcal{A} \times \mathcal{A}$.*

The attack relation is defined on the basis of defeat $\mathcal{D}$ and preference relation $\succeq$, and therefore also the other relations defined by Dung are reused by the preference-based argumentation framework.

Definition 8 *Let $\langle \mathcal{A}, \mathcal{R} \rangle$ be an argumentation framework and $\langle \mathcal{A}, \mathcal{D}, \succeq \rangle$ a preference-based argumentation framework. We say that $\langle \mathcal{A}, \mathcal{D}, \succeq \rangle$ represents $\langle \mathcal{A}, \mathcal{R} \rangle$ iff for all arguments A and B of $\mathcal{A}$, we have $A \; \mathcal{R} \; B$ iff $A \; \mathcal{D} \; B$ and it is not the case that $B \succ A$. We also say that $\mathcal{R}$ is represented by $\mathcal{D}$ and $\succeq$.*

From this definition follows immediately that when $\succeq$ is a total pre-order, we have: $A \mathcal{R} B$ iff $A \mathcal{D} B$ and $A \succeq B$.

Preference reasoning

In most preference-based argumentation frameworks, the preference order on arguments is based on an evaluation of single arguments (Amgoud, Cayrol, & LeBerre 1996). It consists in computing the strength of the argument on the basis of knowledge from which it is built, knowledge being pervaded with implicit or explicit priorities. Note however that knowledge is not always pervaded with priorities which makes it difficult to use this way to evaluate arguments. Moreover one may also need to express more sophisticated preferences such as preferences among *sets* of *abstract* arguments without referring to their internal structure. We adapt in this paper a preference logic of non-monotonic reasoning (Kaci & van der Torre 2005) to the context of argumentation framework. Let p and q be two values. A preference of p over q, denoted $p >> q$, is interpreted as a preference of arguments promoting p over arguments promoting q.

Definition 9 (Value based argumentation framework) *A value based argumentation framework (VAF) is a 5-tuple $\langle \mathcal{A}, \mathcal{D}, V, >>, arg \rangle$ where $\mathcal{A}$ is a set of arguments, $\mathcal{D}$ is a defeat relation, V is a set of values, $>>$ is a total or partial order on V, called a preference specification, and arg is a function from V to $2^{\mathcal{A}}$ s.t. $arg(v)$ is the set of arguments supporting the value v.*

Given a preference specification the logic allows to compute a total pre-order over the set of all arguments. We are interested here in computing a *unique* total pre-order that satisfies the preference specification. Let $\succeq$ be the total pre-order that we intend to compute. A preference of p over q may be interpreted in two ways:

(1) either we compare the *best* arguments in favor of p and the *best* arguments in favor of q w.r.t. $\succeq$. In this case we say that $\succeq$ satisfies $p >> q$ iff $\forall A \in \max(arg(p), \succeq), \forall B \in \max(arg(q), \succeq)$ we have $A \succ B$.

(2) or we compare the *worst* arguments in favor of p and the *worst* arguments in favor of q w.r.t. $\succeq$. In this case we say that $\succeq$ satisfies $p >> q$ iff $\forall A \in \min(arg(p), \succeq), \forall B \in \min(arg(q), \succeq)$ we have $A \succ B$.

Comparing the worst arguments of $arg(p)$ and the best arguments of $arg(q)$ w.r.t. $\succeq$ can be reduced to comparing single arguments (see the related work). So they can be used in both above items. Comparing the best arguments of $arg(p)$ and the worst arguments of $arg(q)$ w.r.t. $\succeq$ is irrelevant (Kaci & van der Torre 2005).

Definition 10 (Model of a preference specification) *$\succeq$ satisfies (or is a model of) a preference specification $\mathcal{P} = \{p_i >> q_i : i = 1, \cdots, n\}$ iff $\succeq$ satisfies each $p_i >> q_i$ in $\mathcal{P}$.*

The above two cases correspond to two different reasonings: an *optimistic* reasoning which applies to the first case since we compare the best arguments w.r.t. $\succeq$, and a *pessimistic* reasoning which applies to the second case since we compare the worst arguments w.r.t. $\succeq$.

The optimistic reasoning corresponds to the minimal specificity principle in non-monotonic reasoning (Pearl 1990). Following this principle there is a unique model of $\mathcal{P}$. This model, called the least specific model of $\mathcal{P}$, is characterized as gravitating towards the ideal since arguments are put in the highest possible rank in the pre-order $\succeq$. The pessimistic reasoning behaves in an opposite way and corresponds to the maximal specificity principle in non-monotonic reasoning. Following this principle there is also a unique model of $\mathcal{P}$ (Benferhat *et al.* 2002). This pre-order, called the most specific model of $\mathcal{P}$, is characterized as gravitating towards the worst since arguments are put in the lowest possible rank in the pre-order $\succeq$.

Definition 11 (Minimal/Maximal specificity principle) *Let $\succeq$ and $\succeq'$ be two total pre-orders on a set of arguments $\mathcal{A}$ represented by ordered partitions $(E_1, \cdots, E_n)$ and $(E'_1, \cdots, E'_m)$ respectively. We say that $\succeq$ is at least as specific as $\succeq'$, written as $\succeq \sqsubseteq \succeq'$, iff $\forall A \in \mathcal{A}$, if $A \in E_i$ and $A \in E'_j$ then $i \leq j$.*
$\succeq$ belongs to the set of the least (resp. most) specific pre-orders among a set of pre-orders $\mathcal{O}$ if there is no $\succeq'$ in $\mathcal{O}$ such that $\succeq' \sqsubset \succeq$, i.e., $\succeq' \sqsubseteq \succeq$ holds but $\succeq \sqsubseteq \succeq'$ (resp. $\succeq \sqsubset \succeq'$) does not.

Since the preference-based argumentation framework is mainly based on the preference relation among arguments, it is worth noticing that the choice of the reasoning attitude is predominant in the output of the argumentation system.

Example 1 *Let $\mathcal{A} = \{A, B, C\}$ be a set of arguments and $V = \{p, q\}$ be the set of values. Let $\mathcal{D}$ be a defeat relation defined by $C \mathcal{D} B$ and $B \mathcal{D} C$. Let $p >> q$ with $arg(p) = \{A\}$ and $arg(q) = \{B\}$. Following the optimistic reasoning the total pre-order satisfying $p >> q$ is $\succeq_o = (\{A, C\}, \{B\})$. We can check that each argument is put in the highest possible rank in $\succeq_o$ s.t. $p >> q$ is satisfied. So we have C attacks B. The grounded extension is composed of A and C. Now following the pessimistic reasoning the total pre-order satisfying $p >> q$ is $\succeq_p = (\{A\}, \{B, C\})$. Here also we can check that each argument is put in the lowest possible rank in $\succeq_p$ s.t. $p >> q$ is satisfied. In this case we have B attacks C and C attacks B. The grounded extension is composed of A only.*

Note that in this example pessimistic reasoning returns less acceptable arguments than optimistic reasoning, however this is not always the case. In addition to the defeat relations given in Example 1 we give $A \mathcal{D} C$ and $C \mathcal{D} A$. Then following the optimistic reasoning the grounded extension is empty while following the pessimistic reasoning the grounded extension is $\{A\}$.
Let us now consider the same example but with the following defeat relations $A \mathcal{D} C$ and $C \mathcal{D} A$ only. Then the grounded extension following the optimistic reasoning is $\{B\}$ while the grounded extension following the pessimistic reasoning is $\{A, B\}$.
Indeed the two kinds of reasoning are incomparable. It is important to notice that the optimistic/pessimistic adjectives refer to the way the arguments are ranked in the total pre-order $\succeq$.

Grounded extension in optimistic reasoning

Algorithms of optimistic reasoning compute the total pre-order $\succeq$ starting from the *best* arguments w.r.t. $\succeq$. Indeed this property makes it possible to compute incrementally the grounded extension when computing this pre-order. Informally this consists in first computing the set of the best arguments w.r.t. $\succeq$. Let us say E_0. Then arguments in E_0 which are not defeated in E_0 belong to the grounded extension. Also all arguments in E_0 defeated only by arguments in $\mathcal{A}\backslash E_0$ belong to the grounded extension. Belong also to the grounded extension arguments in E_0 which are defeated by arguments in E_0 but defended by acceptable arguments, i.e., arguments already put in the grounded extension. Lastly all arguments in $\mathcal{A}\backslash E_0$ defeated by arguments in the current grounded extension will certainly not belong to the grounded extension and can be removed from $\mathcal{A}$. Once $\mathcal{A}$ updated we compute the set of immediately preferred arguments, let's say E_1. At this stage non defeated arguments from E_1 are added to the current grounded extension. Also belong to the grounded extension arguments in E_1 which are defeated by arguments in E_1 but their defeaters are themselves defeated by the current grounded extension. This means that these arguments are defended by the grounded extension. Lastly all arguments in $\mathcal{A}\backslash E_1$ defeated by selected arguments (in the grounded extension) are discarded. This reasoning is repeated until the set of arguments is empty. Algorithm 1 gives a formal description of our procedure to compute progressively the grounded extension. Let

- $Safe(E_l) = \{B : B \in E_l$ s.t. $\nexists B' \in (E_l \cup R)$ with $B'\mathcal{D}B\}$,
- $Acceptable_{\mathcal{GE}}(E_l) = \{B : B \in E_l$ s.t. for each $B' \in (E_l \cup R)$ s.t. $B'\mathcal{D}B, \exists C \in \mathcal{GE}$ s.t. $C\mathcal{D}B'\}$,
- $non\text{-}Safe(\mathcal{A}) = \{B : B \in \mathcal{A}$ s.t. $\exists B' \in \mathcal{GE}$ with $B'\mathcal{D}B\}$.

Algorithm 1: Computing the grounded extension in optimistic reasoning.

Data: $\langle \mathcal{A}, \mathcal{D}, V, >>, arg\rangle$.
Result: The grounded extension.
begin
$l = 0, \mathcal{GE} = \emptyset, R = \emptyset$;
while $\mathcal{A} \neq \emptyset$ **do**
- $E_l = \{B : B \in \mathcal{A}, \forall p_i >> q_i, B \notin arg(q_i)\}$;
if $E_l = \emptyset$ **then** Stop (inconsistent preferences);
– $\mathcal{GE} = \mathcal{GE} \cup Safe(E_l)$;
– $\mathcal{GE} = \mathcal{GE} \cup Acceptable_{\mathcal{GE}}(E_l)$;
– $\mathcal{A} = \mathcal{A}\backslash E_l$;
– $\mathcal{A} = \mathcal{A}\backslash\ non\text{-}Safe(\mathcal{A})$;
– $R = R \cup (E_l \backslash \mathcal{GE})$;
/** remove satisfied preferences **/;
– remove $p_i >> q_i$ where $arg(p_i) \cap E_l \neq \emptyset$;
– $l = l + 1$.
return $\mathcal{GE}$
end

Example 2 *Tom and Mary discuss with their children about their education. Several arguments are given concerning the plans of the children to spend the day. In an attempt to structure the discussion, the arguments are grouped according to several values they promote, and modeled as follows. Tom and Marry give the following set of preferences* $\{Health >> Unhealth, Education >> Enjoy, Social >> Alone\}$.
Let $\mathcal{A} = \{A_0, A_1, A_2, A_3, A_4, A_5, A_6, A_7\}$ *be a set of arguments where* $arg(Health) = \{A_4, A_5\}$, $arg(Unhealth) = \{A_6, A_7\}$, $arg(Education) = \{A_3, A_5, A_7\}$, $arg(Enjoy) = \{A_2, A_4, A_6\}$, $arg(Social) = \{A_0, A_4\}$ *and* $arg(Alone) = \{A_1, A_5\}$.
Let the following defeat relations $A_6\mathcal{D}A_0$, $A_0\mathcal{D}A_6$, $A_3\mathcal{D}A_4$, $A_3\mathcal{D}A_2$, $A_2\mathcal{D}A_5$, $A_5\mathcal{D}A_2$, $A_4\mathcal{D}A_5$ *and* $A_5\mathcal{D}A_4$. *Figure 1 summarizes defeat relations among the arguments. An arrow from A to B stands for "A defeats B".*

We first put in E_0 *arguments which are not in*

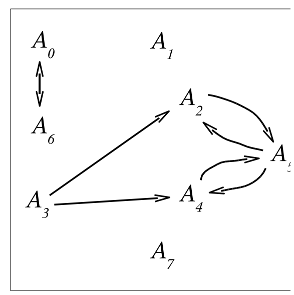

Figure 1: Defeat relations among the arguments.

$arg(Unhealth)$, $arg(Enjoy)$ *and* $arg(Alone)$. *We get* $E_0 = \{A_0, A_3\}$. R *is the empty set and there is no defeat relation among arguments of* E_0 *so both* A_0 *and* A_3 *are safe. They belong to* $\mathcal{GE}$. *The set* $Acceptable_{\mathcal{GE}}(E_0)$ *returns the empty set since there is no defeat relation in* E_0. *Now we first remove arguments of* E_0 *from* $\mathcal{A}$ *since they have been treated. We get* $\mathcal{A} = \{A_1, A_2, A_4, A_5, A_6, A_7\}$. *Then we remove from* $\mathcal{A}$ *arguments which are defeated by arguments in* $\mathcal{GE}$ *(i.e. which are already accepted). We remove* A_2, A_4 *and* A_6. *So* $\mathcal{A} = \{A_1, A_5, A_7\}$. $R = \emptyset$ *since* $E_0 = \mathcal{GE}$. *Lastly we remove* $Education >> Enjoy$ *and* $Social >> Alone$ *since they are satisfied. We run the second iteration of the algorithm. We have* $E_1 = \{A_1, A_5\}$. A_1 *and* A_5 *are safe so they are added to* $\mathcal{GE}$, *i.e.,* $\mathcal{GE} = \{A_0, A_3, A_1, A_5\}$. $Acceptable_{\mathcal{GE}}(E_1)$ *is empty. We remove* A_1 *and* A_5 *from* $\mathcal{A}$. *There are no non-safe arguments in* $\mathcal{A}$ *and* $R = \emptyset$. *In the third iteration of the algorithm we have* $E_2 = \{A_7\}$. A_7 *is safe so* $\mathcal{GE} = \{A_0, A_3, A_1, A_5, A_7\}$.

The role of the set R does not appear in this example however it is important to define such a set to compute incrementally the grounded extension. Let $\mathcal{A} = \{A, B, C, D\}$ be a set of arguments such that $B\mathcal{D}C$, $C\mathcal{D}B$ and $B\mathcal{D}D$. Suppose that the first iteration gives $E_0 = \{A, B, C\}$. So A belongs to the grounded extension while B and C do not (since they attack each other). Following the algorithm we update $\mathcal{A}$ and get $\mathcal{A} = \{D\}$. At this stage it is important to keep B and C in a set, let's say R. The reason is that in the second iteration of the algorithm we should not put D in the grounded extension just because it is not defeated by A. In fact D is attacked by B and not defended by A. This justifies why we consider $E_l \cup R$ when computing $Safe(E_l)$ and $Acceptable_{\mathcal{GE}}(E_l)$.

Let us now first compute the pre-order and then compute the grounded extension. We compute this pre-order from Algorithm 1 by replacing *while* loop by

while $\mathcal{A} \neq \emptyset$ **do**

- $E_l = \{B : B \in \mathcal{A}, \forall p_i >> q_i, B \notin arg(q_i)\}$;
- remove $p_i >> q_i$ where $arg(p_i) \cap E_l \neq \emptyset$.

We have $\succeq_o = (E_0, E_1, E_2)$ where $E_0 = \{A_0, A_3\}$, $E_1 = \{A_1, A_2, A_4, A_5\}$ and $E_2 = \{A_6, A_7\}$.
Let us now compute the grounded extension. Following Definition 8 the attack relations are $A_3\mathcal{R}A_2$, $A_3\mathcal{R}A_4$, $A_4\mathcal{R}A_5$, $A_5\mathcal{R}A_4$, $A_2\mathcal{R}A_5$, $A_5\mathcal{R}A_2$ and $A_0\mathcal{R}A_6$. We first put in the grounded extension arguments which are not attacked, so $\mathcal{GE} = \{A_0, A_1, A_3, A_7\}$. Then we add to $\mathcal{GE}$ arguments which are attacked but defended by arguments in $\mathcal{GE}$. We add A_5. So $\mathcal{GE} = \{A_0, A_1, A_3, A_7, A_5\}$.

The following theorem shows that Algorithm 1 computes the grounded extension.

Theorem 1 *Let $\mathcal{F} = \langle \mathcal{A}, \mathcal{D}, V, >>, arg\rangle$ be a VAF. Algorithm 1 computes the grounded extension of $\mathcal{F}$.*

Grounded extension in pessimistic reasoning

A particularity of pessimistic reasoning is that it computes the total pre-order starting from the lowest ranked arguments in this pre-order. Indeed it is no longer possible to compute progressively the grounded extension. Let us consider our running example. Following the pessimistic reasoning (we will give the formal algorithm later in this section), the worst arguments are A_1, A_2 and A_6. At this stage we can only conclude that A_1 belongs to $\mathcal{GE}$ since it is not defeated. However the status of A_2 and A_6 cannot be determined since they are attacked by A_3 and A_0 respectively. Since higher ranks in $\succeq$ are not computed yet we cannot check whether A_3 and A_0 are attacked or not. The only case where the status of A_2 and A_6 can be determined is when at least one of their defeaters is not defeated. In this case we can conclude that they do not belong to $\mathcal{GE}$. Algorithm 2 gives the total pre-order following the pessimistic reasoning. Each argument is put in the lowest possible rank in the computed pre-order.

Example 3 *(cont'd)*
We put in E_0 arguments which do not appear in any $arg(Health)$, $arg(Education)$ and $arg(Social)$. We get

Algorithm 2: Pessimistic reasoning.

Data: $\langle \mathcal{A}, \mathcal{D}, V, >>, arg\rangle$.
Result: A total pre-order $\succeq_p$ on $\mathcal{A}$.
begin
  $l = 0$;
  while $\mathcal{A} \neq \emptyset$ **do**
    $E_l = \{B : B \in \mathcal{A}, \forall p_i >> q_i, B \notin arg(p_i)\}$;
    if $E_l = \emptyset$ **then** Stop (inconsistent preferences);
    – Remove from $\mathcal{A}$ elements of E_l;
    /** remove satisfied preferences **/
    – Remove $p_i >> q_i$ where $arg(q_i) \cap E_l \neq \emptyset$;
    – $l = l + 1$.
  return $(E'_1, \cdots, E'_{l-1})$ *s.t.* $\forall 1 \leq h \leq l, E'_h = E_{l-h-1}$
end

$E_0 = \{A_1, A_2, A_6\}$. *We remove all preferences $p_i >> q_i$ s.t. $arg(q_i) \cap E_0 \neq \emptyset$. All preferences are removed. Then* $E_1 = \{A_0, A_3, A_4, A_5, A_7\}$.
So we have $\succeq_p = (\{A_0, A_3, A_4, A_5, A_7\}, \{A_1, A_2, A_6\})$.

In this example we get the same grounded extension as in the optimistic reasoning. However if we add for example the defeat relations $A_3\mathcal{D}A_7$ and $A_7\mathcal{D}A_3$ then the grounded extension following the optimistic reasoning is $\{A_0, A_1, A_3, A_5\}$ while following the pessimistic reasoning the grounded extension is $\{A_0, A_1\}$.

Related Work

The preference-based argumentation theory introduced in this paper integrates several existing approaches, most notable the preference based framework of Amgoud and Cayrol (Amgoud & Cayrol 2002), and the value based argumentation theory of Bench-Capon (Bench-Capon 2003). However, there are also substantial conceptual and technical distinctions.

Maybe the main conceptual distinction is that the above authors present their argumentation theory as an extension of Dung's framework, which has the technical consequence that they also define new notions of, for example, defence and acceptance. We, in contrast, consider our preference-based argumentation theory as an alternative representation of Dung's theory, that is, as a kind of front end to it, which has the technical consequence that we do not have to introduce such notions. Reductions of the other preference-based argumentation theories to Dung's theory may be derived from some of the results presented by these authors.

Another conceptual distinction is that in our theory, there seems to be a higher impact of preference reasoning in argumentation. The preference ordering on arguments is not given, but has to be derived from a more abstract preference specification. Technically, this leads to our use of non-monotonic preference reasoning to derive a Dung-style attack relation from a preference specification together with a defeat relation. None of the existing approaches studies the use of non-monotonic reasoning techniques to reason with the preferences. Another conceptual distinction with the work of Bench-Capon is that he, following Perelman, is concerned with an audience.

Concerning the extensive work of Amgoud and colleagues on preference-based argumentation theory, our preference based argumentation theory seems closest to the argumentation framework based on contextual preferences of Amgoud, Parsons and Perrussel (Amgoud, Parsons, & Perrussel 2000). A context may be an agent, a criterion, a viewpoint, *etc.*, and they are ordered. For example, in law earlier arguments are preferred to later ones, arguments of a higher authority are preferred to arguments of a lower authority, more specific arguments are preferred over more general arguments, and these three rules are ordered themselves too. However our approach is more general since we compare sets of arguments instead of single arguments as it is the case in their approach. Bench-Capon (Bench-Capon 2003) develops a value-based argumentation framework, where arguments promote some value. No ordering is required among arguments promoting the same value. If a value V is prioritized over another value W then this is interpreted as "each argument promoting the value V is preferred to all arguments promoting the value W". In our framework we can add such preferences, or encode them as $p_i >> q_j$ where p_i is an argument in favor of V and q_j is an argument in favor of W. Note that in our example there is no ordering which satisfies these strong preferences.

Specificity principle we used in this paper has been also used in many other works (Prakken & Sartor 1997; Poole 1985; Simari & Loui 1992; Stolzenburg *et al.* 2003)[1] however in that works preference relation over arguments is defined on the basis of specificity of their internal structure. In fact arguments are built from default and strict knowledge. Then an argument is preferred to another if its internal structure is more specific. In our work specificity concerns *abstract* arguments without referring to their internal structure.

There are numerous works on non-monotonic logic and in particular the non-monotonic logic of preference which is related to the work in this paper, and which can be used to further generalize the reasoning about preferences in argumentation. Interestingly, as argumentation theory is itself a framework of non-monotonic reasoning, due to our non-monotonic reasoning about preferences two kinds of non-monotonicity seems to be present in our system; we leave a further analysis of this phenomena for further research.

[1] Note that Poole (Poole 1985) uses specificity of arguments without studying interaction among arguments.

Summary

To promote a higher impact of preference reasoning in argumentation, we introduce a novel preference-based argumentation theory. Starting with a set of arguments, a defeat relation, and an ordered set of values, we use the ordered values to compute a preference relation over arguments, and we combine this preference relation with the defeat relation to compute Dung's attack relation. Then we use any of Dung's semantics to define the acceptable set of arguments. In contrast to most other approaches, our approach to reason about preferences in argumentation does not refer to the internal structure of the arguments.

The problem of reducing ordered values to a preference relation comes down to reducing a preference relation over sets of arguments to a preference relation over single arguments. To reason about ordered values and to compute the preference relation over arguments, we are inspired by insights from the non-monotonic logic of preference known as minimal specificity, System Z, gravitation to normality, and otherwise, and we use both so-called optimistic and pessimistic ways to define the preference relation.

For the combined approach, we introduce an algorithm for Dung's grounded semantics. It shows that the computation of the set of acceptable arguments can be combined with the optimistic reasoning to incrementally define the set of acceptable arguments, because in this construction for each equivalence class we can deduce which arguments are not attacked by other arguments. This property does not hold for pessimistic reasoning.

In future work, we study other ways to use reasoning about preferences in argumentation theory. For example, Bochman (2005) develops a generalization of Dung's theory, called collective argumentation, where the attack relation is defined over sets of arguments instead of single arguments. It seems natural to develop a unified framework where both attack and preference relations are defined over sets of arguments. Another future work is to study the reinforcement among different arguments promoting the same value as advocated in (Bench-Capon 2003).

References

Amgoud, L., and Cayrol, C. 2002. A reasoning model based on the production of acceptable arguments. *AMAI Journal* 34:197–216.

Amgoud, L.; Cayrol, C.; and LeBerre, D. 1996. Comparing arguments using preference orderings for argument-based reasoning. In *8th International Conf. on Tools with Artificial Intelligence*, 400–403.

Amgoud, L.; Parsons, S.; and Perrussel, L. 2000. An argumentation framework based on contextual preferences. Technical report, Department of Electronic Engineering, Queen Mary and Westfield College.

Atkinson, K.; Bench-Capon, T.; and McBurney, P. 2005. Persuasive Political Argument. In *Computational Models of Natural Argument*, 44–51.

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

Benferhat, S.; Dubois, D.; Kaci, S.; and Prade, H. 2002. Bipolar possibilistic representations. In *18th International Conference on Uncertainty in Artifcial Intelligence (UAI'02)*, 45–52.

Bochman, A. 2005. Propositional Argumentation and

Causal Reasoning. In *11th Int. Joint Conf. on Artificial Intelligence*, 388–393.

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

Kaci, S., and van der Torre, L. 2005. Algorithms for a Nonmonotonic Logic of Preferences. In *8th Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 281–292.

Kaci, S.; van der Torre, L.; and Weydert, E. 2006. Acyclic argumentation: Attack = conflict + preference. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06)*, to appear.

Pearl, J. 1990. System z: A natural ordering of defaults with tractable applications to default reasoning. In *3rd Conference on Theoretical Aspects of Reasoning about Knowledge (TARK'90)*, 121–135.

Poole, D. L. 1985. On the comparison of theories: Preferring the most specific explanation. In *Proceedings of the 9th IJCAI*, 144–147. IJCAI.

Prakken, H., and Sartor, G. 1997. Argument-based extended logic programming with defeasible priorties. *Journal of Applied Non-Classical Logics* 7:25–75.

Simari, G. R., and Loui, R. P. 1992. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence* 53:125–157.

Stolzenburg, F.; García, A. J.; Chesñevar, C. I.; and Simari, G. R. 2003. Computing generalized specificity. *Journal of Aplied Non-Classical Logics* 13(1):87–113.