An abstract argumentation framework with proper and blocking defeaters

Diego C. Martínez   Alejandro García   Guillermo Simari
Laboratorio de Investigación y Desarrollo en Inteligencia Artificial
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur
Av. Alem 1253 - (8000) Bahía Blanca - Buenos Aires - República Argentina
Tel/Fax: (+54)(291)4595135/5136 - E-mail: **{dcm,ajg,grs}@cs.uns.edu.ar**

Abstract

Defeat between arguments is a combination of two basic elements: a conflict relation and a preference order. We present a framework for argumentation where two kinds of defeat relation are present, depending on the outcome of a preference function σ. An argument A is a *proper defeater* of B, if both arguments are in conflict, and A is preferred to B according to σ. A is a *blocking defeater* of B, if they are in conflict and both arguments are *incomparable* or *indistinguishable*. We define a function to caracterize the set of accepted argument in the framework. This function also provide a method for identifying controversial situations.

1 Introduction

In formal systems of defeasible argumentation, arguments for and against a proposition are produced and evaluated to verify the acceptability of that proposition. The main idea in these systems is that any proposition will be accepted as true if there exists an argument that supports it, and this argument is acceptable according to an analysis between it and its counterarguments. This analysis requires a process of comparison of conflicting arguments, in order to decide which one is preferable. After this dialectical analysis over the set of arguments in the system, some of them will be *acceptable* or *justified* arguments, while others not. Argumentation is widely used in nonmonotonic reasoning [6, 3] and it is suitable to model dialogues between intelligent agents [1, 5].

Abstract argumentation systems [8, 2, 4] are formalisms for argumentation, where some components remains unspecified, usually the structure of arguments. In this kind of systems, the emphasis is put on semantic notions, basically the task of finding the set of accepted arguments. Most of them are based on a single abstract notion called *attack relation*, and several argument extensions are defined as sets of possible accepted arguments. However, the task of comparing arguments to establish a preference is not always successful. Finding a preferred argument is essential to determine a defeat relation.

In the next section, we present an abstract framework for argumentation where conflicts and preference between arguments are considered. These two elements allows the definition of two kind of argument defeat relations.

2 Argumentation Framework

Our argumentation framework is formed by three elements: a set of arguments, a binary conflict relation over this set, and some function used to evaluate the relative difference of conclusive force for any pair of arguments.

Definition 2.1. *An argumentation framework AF is the triplet* $< Args, \mathcal{C}, \sigma >$ *where Args is a set of arguments,* $\mathcal{C}$ *is a binary conflict relation between arguments,* $\mathcal{C} \subseteq Args \times Args$ *and* $\sigma : Args \times Args \rightarrow 2^{Args}$ *is a preference function for conflictive arguments.*

Arguments are abstract entities, as in [2], denoted by uppercase letters. If A is an argument, then A^- is a subargument of A, and A^+ is a superargument of A. No reference to the underlying logic is needed. It is sufficient to know that arguments support conclusions, which are denoted here by lowercase letters, when needed.

Example 2.1. *The following are argumentation frameworks. Only relevant cases of function* σ *are shown, those involving conflicting arguments.*

- $AF_1 =< Args, \mathcal{C}, \sigma >$ *where* $Args = \{A, B, C\}$ *and* $\mathcal{C} = \{(A, B), (B, C)\}$ *and* $\sigma(A, B) = \{A\}$ *and* $\sigma(B, C) = \{C\}$.

- $AF_2 =< Args, \mathcal{C}, \sigma >$ *where* $Args = \{A, B, C, D\}$ *and* $\mathcal{C} = \{(A, B), (C, D)\}$ *and* $\sigma(X, Y) = \{\}$ *for all X and Y.*

- $AF_3 =< Args, \mathcal{C}, \sigma >$ *where* $Args = \{A, B, C, D\}$ *and* $\mathcal{C} = \{(A, B), (B, C), (C, D)\}$ *and* $\sigma(A, B) = \{A\}, \sigma(B, C) = \{B\}$ *and* $\sigma(C, D) = \{C, D\}$.

The conflict relation between two arguments A and B denotes the fact that these arguments can not be accepted simultaneously, because they contradict each other. For example, two arguments A and B that support complementary conclusions h and $\neg h$ can not be accepted together. Conflict relations are denoted by unordered pairs of arguments, and the set of all conflict relations on AF is denoted by $\mathcal{C}$. Given a set of arguments T, an argument $A \in S$ is said to be in conflict in S, if there is an argument $B \in S$ such that $(A, B) \in \mathcal{C}$. For any argument A, the set of all arguments in conflict with A is denoted by $Conf(A)$.

The constraints imposed by the conflict relation lead to several sets of possible accepted arguments. For example, if $Args = \{A, B\}$ and $(A, B) \in \mathcal{C}$, then $\{A\}$ is a set of possible accepted arguments, and so is $\{B\}$. Therefore, a clear decision must be taken. In order to accomplish this task, function σ is used to evaluate arguments, comparing them to establish a preference based on the conclusive force.

Definition 2.2. *An argument comparison criterion is a function* $\sigma : S \times S \to 2^S$*, where* S *is the set of arguments in the framework and* $\sigma(A, B) \subseteq \mathcal{P}(\{A, B\})$*. If* $\sigma(A, B) = \{A\}$ *then* A *is preferred to* B*. In the same way, if* $\sigma(A, B) = \{B\}$ *then* B *is preferred to* A*. If* $\sigma(A, B) = \{A, B\}$ *then* A *and* B *are arguments with equal relative strength. If* $\sigma(A, B) = \emptyset$ *then* A *and* B *are incomparable arguments.*

For two arguments A and B, such that $(A, B) \in \mathcal{C}$ there are four possible outcomes, according to definition 2.2:

- $\sigma(A, B) = \{A\}$. In this case a *defeat* relation is established. Because A is preferred to B, in order to accept B it is necessary to analyze the acceptance of A, but not the other way around. It is said that argument A *defeats* argument B, and A is a *proper defeater* of B.

- $\sigma(A, B) = \{B\}$. In a similar way, argument B *defeats* argument A, and therefore B is a *proper defeater* of A.

- $\sigma(A, B) = \{A, B\}$. Both arguments are equivalent, i.e, there is no relative difference of conclusive force, so A and B are said to be *indistinguishable*. No proper defeat relation can be established between these arguments.

- $\sigma(A, B) = \emptyset$. Both arguments are *incomparable* according to σ, and no *proper* defeat relation is inferred.

In the first two cases, a concrete preference is made between two arguments, and therefore a defeat relation is established. The preferred arguments are called *proper defeaters*. In the last two cases, no preference is made, either because both arguments are indistinguishable to each other or they are incomparable. The conflict between these two arguments remains unsolved. Due to the fact that the conflict relation is a symmetric relation, an argument *blocks* the other one and it is said that both of them are *blocking defeaters* [6, 7]. An argument B is said to be a $defeater$ of an argument A, if B is a blocking or a proper defeater of A. In example 2.1, in the argumentation framework AF_3, argument A is a proper defeater of argument B while C is a blocking defeater of D and viceversa.

3 Semantic of Argumentation

Popular semantics on argumentation frameworks are based on defeat relations, usually called *attack* relations, as in [2]. These formalisms assume the existence of a binary relation of attack, such that when (A, B) are in relation then in order to accept B it is necessary to find out if A is not accepted, but not the other way around. Few authors remark that this asymmetric relation should be derived from a conflict relation between arguments and a suitable comparison criterion, not specified in their abstract systems. However, as stated in our framework, the comparison method used to evaluate pairs of arguments may not always establish a preference on conflictive arguments. In fact, a method which is able to determine the relative difference in strength between any pair of arguments in the system is not a very realistic one. There are, by nature, incomparable arguments [8]. Therefore, another special kind of defeat must be taken into account: the one derived when no preference can be established between conflictive arguments.

We believe an extended semantic for argumentation is needed. This semantic is based on the conflict relation between arguments and the comparison criterion used to evaluate these arguments, rather than a single abstract defeat relation.

Some common sense premises must be observed:

1. The status of any argument A is determined by the status of arguments in conflict with A.

2. Any pair of conflictive arguments may be *decided* or *undecided*. A pair (A, B) is decided when $|\sigma(A, B)| = 1$, and it is undecided on the contrary.

For two conflictive arguments A and B, there seems to be no difference between being incomparable and being equivalent in conclusive force. In the former case, there is no defeat relation on A and B. In the latter, as the arguments are indistinguishable, usually they are said to be *mutual defeaters*. Despite this difference, in both cases the dialectical implications are the same: argument A may be accepted only if B is not accepted, and viceversa.

Any bi-valued semantic for argumentation will classify the arguments involved in an undecided conflict relation as non-accepted arguments, unless one of them is defeated by an accepted argument. In [2] several examples are shown in which circular argumentation leads to an empty set of accepted arguments. Jakobovits proposes in [4] a three-valued classification of arguments, where arguments involved in an undecided conflict relation are candidates to be labeled as *undecided arguments*. However, these systems are build upon classical attack relation.

We want a bi-valued semantic to deal with proper and blocking defeat relations. No argument extensions are considered [2], as any argument in the system is classified as an accepted or non-accepted argument. Even more, the propagation of indecision due to incomparable arguments (or equivalent in conclusive force) is limited in some cases, based on reasonable dialectical intepretations [3].

3.1 Classifying arguments

The notion of conflict between two arguments A and B establishes that these arguments can not be accepted together. Therefore, any set of accepted arguments must not include conflictive arguments. This basic concept is formalized in the next definition.

Definition 3.1. *A set of arguments S is coherent, if for all $A \in S$ such that $(A, B) \in \mathcal{C}$, then $B \notin S$. A coherent set of arguments is said to be maximal, if there is no coherent set S' such that $S \subseteq S'$.*

Coherent sets of arguments corresponds to conflict free sets in Dung's semantics [2], where "*conflict*" actually means "*attack*", a classical form of defeat.

Example 3.1. *Let $AF =< Args, \mathcal{C}, \sigma >$ be an argumentation framework where $Args = \{A, B, C\}$ and $\mathcal{C} = \{(A, B)(B, C)\}$. Then $\{A\}$, $\{B\}$, $\{C\}$ and $\{A, C\}$ are coherent sets of arguments.*

Note that if $(A, B) \in \mathcal{C}$ then there are at least two coherent sets in the framework, say S_A and S_B, the former including A (and not B), and the latter including B (but not A). It is also important to note that any maximal coherent set must include every non-conflictive argument, that is, every argument A such that $(A, B) \notin \mathcal{C}$ for any argument B.

Definition 3.2. *Two coherent sets of arguments S_1 and S_2 are contradictory, if there is an argument $A \in S_1$ and an argument $B \in S_2$ such that $(A, B) \in \mathcal{C}$. The pair (A, B) is called a conflict point.*

Example 3.2. *In example 3.1 coherent sets $\{A, C\}$ and $\{B\}$ are contradictory. The same is true for $\{B\}$ and $\{C\}$.*

Any conflict involving an argument A determines two set of maximal coherent sets, ${\Gamma_A}^+$ and ${\Gamma_A}^-$. All maximal coherent sets that includes A are in ${\Gamma_A}^+$, and any other set is in ${\Gamma_A}^-$.

Proposition 3.1. *For any argument A, $Conf(A) \cap {\Gamma_A}^+ = \emptyset$*

When two arguments are involved in a conflict relation, they are included in different coherent sets. For any set S, it is possible to construct another set including arguments in conflict with elements of S.

Definition 3.3. *Two coherent sets of arguments S_1 and S_2 are complementary, if for every argument $A \in S_1$ there exists an argument $B \in S_2$ such that $(A, B) \in \mathcal{C}$. In other words, every argument in S_1 and S_2 is taking part of a conflict point.*

Example 3.3. *Let $AF =< Args, \mathcal{C}, \sigma >$ be an argumentation framework where $Args = \{A, B, C\}$ and $\mathcal{C} = \{(A, B)(B, C)\}$. Then $\{A, C\}$ and $\{B\}$ are complementary sets of arguments, because (A, B) and (B, C) are in $\mathcal{C}$.*

Arguments can be classified as *accepted* arguments or *non-accepted* or *rejected* arguments, according to its context in the framework. Given a set of arguments S, two kinds of arguments are easily identified as accepted arguments. First, those arguments not involved in any conflict in S. Second, those arguments actually involved in a conflict, but preferred to its conflictive arguments, according to function σ. These special arguments are called *defeater free* arguments.

Definition 3.4. *Let S be a set of arguments. An argument A is defeater-free in S, if no argument in S is a defeater of A.*

Defeater-free arguments must be accepted, as no contradictory information is provided in the framework. Note that this classification is contextual to the set in which the argument is included.

As stated before, the semantic of conflict relation states that when an argument A is accepted, any argument in $Conf(A)$ should be rejected. The next definition captures a subset of arguments that should be rejected in the framework. They are called $s$uppressed arguments.

Definition 3.5. *Let S be a set of arguments. An argument $A \in S$ is said to be suppressed in S, if*

1. *there is a defeater-free argument $B \in S$ such that B is a proper defeater of A, or*

2. *the previous condition does not hold, but there is a blocking defeater B of A, and there is no other argument C in S such that C is a blocking defeater of B.*

The first case is trivial: any conflictive arguments must be suppresed when its counterpart in the conflict is accepted. The second case seems to be not obvious, but reflects the case in which arguments taking part of an undecided conflict relations may also be suppressed, when no defense is involved.

Definitions 3.4 and 3.5 are clearly related. Given a set S of arguments, is as easy to identify obviously suppressed arguments as it is to identify inevitably accepted ones. The following function caracterizes the set of arguments not directly suppressed in a given set.

Definition 3.6. *Let S be a set of arguments. The function $\mathcal{N}$ is defined as follows:*

- $\mathcal{N} : 2^{Args} \longrightarrow 2^{Args}$
- $\mathcal{N}(S) = \{A : A \in S \text{ and } A \text{ is not suppressed in } S\}$

For a set of arguments S, $\mathcal{N}(S)$ is the set of non-suppressed arguments in S. It is easy to see that if S is a coherent set of arguments, then $S = \mathcal{N}(S)$. However, the converse is not true, as shown in the next example:

Example 3.4. *Let $< Args, \mathcal{C}, \sigma >$ be an argumentation framework, where $Args = \{A, B, C, D\}$ and $\mathcal{C} = \{(A, B), (B, C), (C, D), (D, A)\}$ and for all arguments X and Y, $\sigma(X, Y) = \emptyset$. Since no arguments in Args is a defeater-free argument, and any argument is involved in at least two conflict relations, then $\mathcal{N}(Args) = Args$.*

By definition, $\mathcal{N}(S)$ includes some or all of the arguments in S. The arguments not included, if any, are those suppressed in S.

Proposition 3.2. *For any set of arguments S, $\mathcal{N}(S) \subseteq S$*

Some arguments in $\mathcal{N}(S)$ may now be suppressed in this set. This is true because some arguments may be classified now as *defeater-free* arguments in $\mathcal{N}(S)$, as its defeaters are suppressed arguments in S. It is possible then to repeatedly apply the function $\mathcal{N}$ to the set of arguments in the framework. Because of proposition 3.2, this process may continue until obtain a set of arguments T such that $T = \mathcal{N}(T)$.

Definition 3.7. *$\mathcal{N}^n$ is defined as: $\mathcal{N}^0$ is $\mathcal{N}(Args)$ and $\mathcal{N}^{n+1} = \mathcal{N} \circ \mathcal{N}^n$. The set of arguments $\mathcal{N}^k, k \geq 0$ such that $\mathcal{N}^k = \mathcal{N}^{k+1}$ is denoted $\mathcal{N}^\omega$.*

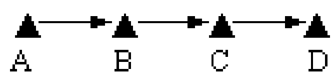

Figure 1: Proper defeaters

Example 3.5. *Let $< Args, \mathcal{C}, \sigma >$ be an argumentation framework, where $Args = \{A, B, C, D\}, \mathcal{C} = \{(A, B), (B, C), (C, D)\}$ and $\sigma(A, B) = \{A\}, \sigma(B, C) = \{B\}$ and $\sigma(C, D) = \{C\}$. Figure 1 shows a graphical representation of this framework, where an arrow from argument X to argument Y means "X is a proper defeater of Y" In this framework,*

$$\mathcal{N}^0 = \{A, C, D\},$$
$$\mathcal{N}^1 = \mathcal{N}(\{A, C, D\}) = \{A, C\}, \text{ and}$$
$$\mathcal{N}^2 = \mathcal{N}(\{A, C\}) = \{A, C\}.$$

Therefore, $\mathcal{N}^\omega = \{A, C\}$

Example 3.6. *Let $< Args, \mathcal{C}, \sigma >$ be an argumentation framework, where $Args = \{A, B, C, D, E, F, G\}, \mathcal{C} = \{(A, B), (A, C), (B, D), (C, E), (F, C), (F, G)\}$. Figure 2 shows a graphical representation of this framework, where arrows represents proper defeat and a double-pointed arrow between two arguments X and Y means "X is a blocking defeater of Y" (and viceversa). In this framework, $\mathcal{N}^0 = \{A, D, E, G\}$. Argument F is suppressed because G is its blocking defeater and G is not in conflict with any other argument. B and C are suppressed because its proper defeaters are defeater-free arguments in $Args$. As the set obtained is* coherent, *$\mathcal{N}^1 = \mathcal{N}^0 = \{A, D, E, G\}$. Therefore, $\mathcal{N}^\omega = \{A, D, E, G\}$*

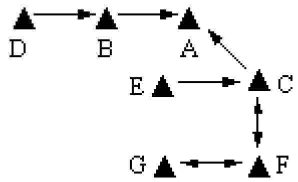

Figure 2: Proper defeaters and blocking defeaters

By definition, no argument is suppressed in $\mathcal{N}^\omega$. Therefore, if $\mathcal{N}^\omega$ is a coherent set (as in examples 3.5 and 3.6), then any argument in $\mathcal{N}^\omega$ is an *accepted* argument.

Definition 3.8. *An argument in $\mathcal{N}^\omega$ such that it is not in conflict with any other argument in the same set is an accepted argument. The set of accepted arguments in $\mathcal{N}^\omega$ is denoted $\mathcal{N}^{\omega+}$.*

As stated before, if $\mathcal{N}^\omega$ is a coherent set of arguments, then $\mathcal{N}^\omega = \mathcal{N}^{\omega+}$. However, as mentioned, this set of arguments may still not be a coherent set. In figure 3, the set $\mathcal{N}^\omega$ is not a coherent set. Even more, no arguments are suppressed in *Args*. This is related to the presence of some special arguments involved in a *fallacy*, as discussed in the next section.

4 Fallacies

When the repeated application of function $\mathcal{N}$ does not lead to a coherent set, then some controversial situations can be found in the argumentation framework. Suppose A_1 and A_2 are two arguments in $\mathcal{N}^\omega$ such that $(A_1, A_2) \in \mathcal{C}$. It is clear that

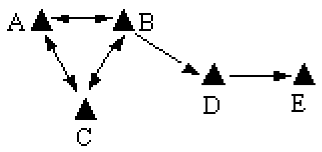

Figure 3: Controversial situation

there is no defeater-free argument C in $\mathcal{N}^\omega$ such that C is a proper defeater of A_1 or A_2. If there exists a proper defeater, then it is not a defeater-free argument. On the other hand, by definition of $\mathcal{N}$, it is possible to state that there is another argument $A_3 \in \mathcal{N}\omega$ such that A_3 is a blocking defeater of A_1 or A_2. Suppose, without loss of generality, that A_3 is a blocking defeater of A_2. As A_2 and A_3 are in $\mathcal{N}^\omega$ then it is possible to state that there is another argument $A_4 \in \mathcal{N}^\omega$ acting as a blocking defeater of A_2 or A_3. The same analysis can be made between A_4 and, say, A_3. As $Args$ is a finite set of arguments, then clearly some $A_i = A_j$, for some i, j such that $i \neq j$. This cycle of defeaters is called a $fallacy$, and the comparison criterion plays a very important role in its existence.

Definition 4.1. *An argumentation framework AF is said to contain a fallacy, if* $\mathcal{N}^\omega$ *is not a coherent set of arguments.*

The presence of fallacies is related to the lack of decision in the preference function, under some common sense conditions, as stated in the next proposition.

Proposition 4.1. *If function* σ *denotes a transitive preference relation, then any fallacy involves one or more arguments acting as mutual blocking defeaters. If, in addition,* σ *denotes a weak order, then fallacies are formed only by indistinguishable arguments, as there are no incomparable arguments in the framework.*

Several preference relations between arguments are used in different argumentation systems. Most of them are based on properties observed on the structure of arguments. Many authors remarks that preference relations must be a total or partial preordering on the set of arguments, and it is widely accepted that a transitive property should be exhibited.

When a cycle includes an even number of defeaters the fallacy is called *circular argumentation*, and *contradictory argumentation* on the contrary. Any argument involved in a fallacy is usually called *fallacious*. The most important premise in defeasible argumentation is that any argument must be *accepted* when none of its defeaters are. However, this fact can not be proved for any fallacious argument, because at least one of its defeaters is also a fallacious argument[1]. Therefore, any argument of this kind should not be accepted.

In figure 3 a simple argumentation framework is depicted. The only five arguments are shown interacting with each other, where a simple arrow means *proper defeat* relation, and a double-pointed arrow means *blocking defeat* relation. Note that, in this framework, $\mathcal{N}^\omega = \mathcal{N}^0$, because no defeater-free arguments are present and there is a cycle of blocking defeaters, so condition 2 from definition 3.5 is not

[1] Any other non-fallacious defeater has been suppressed previously

satisfied. Therefore, $\mathcal{N}^{\omega+} = \emptyset$. Note that not every argument in $\mathcal{N}^{\omega}$ is a fallacious argument: arguments D and E are not involved in the fallacy. However, they are not included in $\mathcal{N}^{\omega+}$ because both of them are directly or indirectly related to fallacious arguments.

5 Conclusions and future work

Defeat between arguments is a combination of two basic elements: a conflict relation and a preference order. Some abstract argumentation systems include these components, but only one kind of defeat is derived or the preference order is used in other purposes. We presented a framework for argumentation where two kinds of defeat relations are present, depending on the outcome of a preference function σ. An argument A is a *proper defeater* of B, if both arguments are in conflict, and A is preferred to B according to σ. On the other hand, A is a *blocking defeater* of B, if they are in conflict and both arguments are *incomparable* or *indistinguishable*. A function to caracterize the set of accepted argument is presented, based on the concept of *suppressed arguments*. The same function can be used to identify argumentation frameworks containing fallacies. Future work will be concerned with establishing a method, based on function $\mathcal{N}$, to deal with controversial situations, in order to allow the acceptance of non-fallacious arguments related directly or indirectly to fallacies.

6 Acknowledgments

We are very grateful to the Department of Computer Science and Engineering, Universidad Nacional del Sur, for making its space and resources available to us at any time. This paper is partially supported by the *Agencia Nacional de Promoción Científica y Tecnológica* (PICT 13096), and the Secretaría de Ciencia y Tecnología of Universidad Nacional del Sur.

The Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA) is a Member of the IICyTI (Instituto de Investigación en Ciencia y Tecnología Informática).

References

[1] Leila Amgoud, Simon Parsons and Nicolas Maudet. *Arguments, dialogue and negotation.* Proceedings of the 14th European Conference on Artificial Intelligence, Berlin, 2000.

[2] Phan Minh Dung. *On the Acceptability of Arguments and its Fundamental Role in Nonmotonic reasoning and Logic Programming and N-persons games*, IJCAI 93 852-857, 1993.

[3] Alejandro García, Carlos Chesñevar, Guillermo Simari. *The Role of Dialectics in Defeasible Argumentation* Proceedings of the XIV International Conference of the Chilean Computer Science Society. Concepción, Chile, Noviembre de 1994.

[4] H. Jakobovitz. *Robust Semantics for argumentation frameworks* Journal of Logic and Computation 9(2):215-261. 1999.

[5] Diego Martinez, Alejandro García *Dialoguing DeLP-based agents.* VIII Congreso Argentino de Ciencias de la Computacion. Facultad de Ciencias Exactas, Universidad Nacional de Buenos Aires. Octubre de 2002.

[6] John Pollock. *Defeasible Reasoning.* Cognitive Science, Vol. 11, pp. 481–518, 1987.

[7] Guillermo R. Simari and Ronald P. Loui. *A Mathematical Treatment of Defeasible Reasoning and its Implementation.* Artificial Intelligence, 53:125–157, 1992.

[8] Gerard A.W. Vreeswijk. *Abstract argumentation systems.* Artificial Intelligence, 90:225–279, 1997.