An abstract argumentation framework with comparison criterion

Diego C. Martínez * Alejandro García
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}@cs.uns.edu.ar**

Abstract

We define an abstract argumentation framework that includes a binary symmetric relation representing argument conflicts, and a function used to evaluate conflicting arguments. In this framework it is clear how defeat relations are constructed, even when the arguments are treated abstractly. We also present a basic classification of comparison criterions and its impact on the set of accepted arguments. Finally, we define the concept of improvements of a comparison criterion and its relation with the fallacies presented in the framework.

1 Introduction

There are a lot of argumentation models that have been developed inside Artificial Intelligence [1, 7, 8]. Among these models, different formal systems of defeasible argumentation are defined, where arguments for and against a proposition are produced and evaluated to verify the acceptability of that proposition. In this manner, defeasible argumentation allows reasoning with incomplete and uncertain information and is suitable to handle inconsistency in knowledge-based systems.

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. Therefore, in the set of arguments of the system, some of them will be *acceptable* or *justified* arguments, while others not.

Almost every system of defeasible argumentation is based on the notion of binary conflicts between arguments [3, 8], and some argument evaluation criterion to

*Becario de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) be applied to conflictive arguments [7]. In the next section we define an abstract argumentation framework where defeat relations are constructed evaluating conflictive arguments

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,* $\mathcal{C} \subseteq Args \times Args$ *and* $\sigma : Args \times Args \rightarrow 2^{Args}$.

Arguments are abstract entities, as in [3], 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.

The most relevant relation in argumentation frameworks is the *defeat* relation (or *attack* relation, as in [3]), built upon the $\mathcal{C}$ relation and function σ. Usually, abstract systems only represent abstract defeat relations. In our framework, the fact that an argument may contradict another argument is represented by the conflict relation $\mathcal{C}$.

2.1 Argument conflicts

The conflict relation between two arguments A and B denotes the fact that these arguments can not be accepted simultaneously, usually because they contradict each other. For example, arguments that support complementary conclusions can not be accepted together. The conflict relation is usually a symmetric one. The most common relation, usually known as the *rebut* relation, is symmetric because an argument is rebutted if there exists an argument for the negated conclusion. That is, an argument A for h is rebutted by an argument B for $\neg h$, and vice versa.

Other well-known relation, called *undercut* relation, is usually presented as a conflict relation. Actually, the *undercut* is the inherited relation from rebutters. It is based on the idea that, if there is a conflict between A and B (they are mutual rebutters), then there is also a conflict between A and B^+, but these arguments do not support complementary conclusions, so they can not be classified as rebutters. Therefore, it is said that argument A *undercuts* argument B^+. Note that this relation is not a symmetric one, but the existence of undercutters depends on the existence of rebutters, so any system including only the rebut relation is implicitly including undercutters. Both conflict relations are used in many argumentation frameworks.

Conflicts are unsolved problems. As stated before, two arguments in conflict can not be accepted together, but no decision is made at this level. What is needed to solve the conflict between arguments is an additional evaluation of the arguments involved. They need to be compared, and the comparison criterion is provided in our framework by function σ.

2.2 Comparing arguments

The main target of the criterion is to identify the relative difference of conclusive force of the arguments being compared.

Definition 2.2. *An argument comparison criterion is a function $\sigma : S \times S \rightarrow 2^S$, where S is the set of arguments in the framework and*

$$\sigma(A, B) = \begin{cases} \{A\} \text{ or} \\ \{B\} \text{ or} \\ \{A, B\} \text{ or} \\ \{\} \end{cases}$$

If $\sigma(A, B) = \{A, B\}$ then A and B are arguments with equal relative strength. If $\sigma(A, B) = \{\}$ then A and B are incomparable arguments.

The comparison criterion takes two arguments A and B and decides which argument is preferred. Popular criteria are usually based on syntax aspects and they just constitute a structural analysis of the involved arguments. Many researchers have proposed general criteria for adjudicating between competing lines of arguments. Most of them agree on the principle of specificity, introduced by Poole [9]. We think that the comparison criterion is one of the most important parts of any argumentation system, as it is responsible of telling how and why a particular argument should overrule any other argument.

The function σ has the following properties:

1. $\sigma(A, B) = \sigma(B, A)$
2. $\sigma(A, A) = \{A, A\}$
3. $\sigma(A, A^-) \supseteq \{A^-\}$
4. If $\sigma(A, B) = \{A\}$ and $\sigma(B, C) = \{B\}$ then $\sigma(A, C) = \{A\}$
5. Monotony: If $\sigma(A, B) = \{A\}$ then $\sigma(A, B^+) = \{A\}$

The first and second properties are obvious. The third property says that an argument is as strong as any subargument. The fourth property is the transitivity property, as stated in [8]. The Monotony property is related to the third property and says that if A is preferred to B, then A is preferred to any superargument of B. Monotony is also related to *self-defense* [1], where an argument is preferred to its undercutter.

Definition 2.3. *An argument comparison criterion is complete if for any pair of arguments A and B, $\sigma(A, B) \neq \{\}$*

A complete criterion is able to determine the difference of conclusive force for any pair of arguments in the framework. No incomparable arguments are found. Complete criteria are based on linear binary relations.

Example 2.1. *The comparison criterion in which the argument with the lowest number of subarguments* [1] *is preferable , is a complete criterion. On the other hand, specificity [7, 9] is not a complete comparison criterion.*

Note that a complete criterion is still able to find arguments with the same relative conclusive force, so in this case no real preference is made.

Definition 2.4. *An argument comparison criterion is $uniform$ if for any pair of arguments A and B, if $A \neq B$ then $\|\sigma(A, B)\| = 1$*

A uniform criterion is always able to find a difference of relative conclusive force for any pair of arguments. This is a desirable property, since the lack of decision in a preference criterion leads to multiple argument extensions.

Proposition 2.1. *Any uniform comparison criterion is complete.*

A uniform criterion has more precision than a complete criterion. In the next section we define how a defeat relation is established and what happens when σ is not able to prefer a particular argument.

2.3 Defeat and controversy

To defeat an argument is to impose a condition to the acceptance of that argument, based on some preference relation. The defeat relation between two conflicting arguments A and B is formally established only when $\|\sigma(A, B)\| = 1$.

Definition 2.5. *Let $AF = (Args, \mathcal{C}, \sigma)$ be an argumentation framework and A and B two arguments in $Args$. An argument A defeats an argument B, denoted $A \xrightarrow{d} B$ if and only if $(A, B) \in \mathcal{C}$ and $\|\sigma(A, B)\| = \{A\}$. In this case, A is called the "defeater argument" and B is called the "defeated argument".*

The relation $A \xrightarrow{d} B$ is a relation of conditional acceptance. It may be interpreted as "in order to accept or reject argument B, we must evaluate the acceptance of A first". Only when a concrete preference is made (that is, $\|\sigma(A, B)\| = 1$) a defeat relation is established. In any other case there is no defeat relation at all. Simply put, the conflict remains unsolved. For this reason, we propose a new kind of relation between arguments called *controversial relation*.

Definition 2.6. *Let $AF = (Args, \mathcal{C}, \sigma)$ be an argumentation framework and A and B two arguments in $Args$. Two arguments A and B are in controversial relation, denoted $A \overset{\times}{\longleftrightarrow} B$ or $B \overset{\times}{\longleftrightarrow} A$, if and only if (A, B)*[2] *$\in \mathcal{C}$ and $\|\sigma(A, B)\| \neq 1$.*

If A and B are involved in a controversial relation, then these arguments are said to be $controversial$. Controversial arguments are either incomparable or they have the same conclusive force. In some systems this kind of relation is usually interpreted as a two-arguments cycle in the defeats graph. This is not very realistic, because no concrete preference was made.

For every argument A in S we define the following sets:

[1] A viariant is based on the number of defeasible steps.

[2] (A, B) is not an ordered pair. The conflict relation is symmetric.

- $I_\sigma(A) = \{B : \sigma(A, B) = \emptyset\}$ is the set of arguments incomparable to A.
- $U_\sigma(A) = \{B : \sigma(A, B) = \{B\}\}$ is the set of arguments preferable to A.
- $D_\sigma(A) = \{B : \sigma(A, B) = \{A\}\}$ is the set of arguments which A is preferable to.
- $E_\sigma(A) = \{B : \sigma(A, B) = \{A, B\}\}$ is the set of arguments equivalent to A in conclusive force.

The candidates for defeat relations containing A (as a defeater or defeated argument) are the arguments in $U_\sigma(A) \cup D_\sigma(A)$. Arguments in $I_\sigma(A) \bigcup E_\sigma(A)$ are candidates for controversial relations on A. Note that if σ is a uniform comparison criterion, then $I_\sigma(A) = E_\sigma(A) = \emptyset$ for any argument A.

Definition 2.7. *Let $AF = (Args, \mathcal{C}, \sigma)$ be an argumentation framework and A $Args$. The following sets are defined:*

- $\mathcal{D}^-(A) = \{B : B \in Args, B \xrightarrow{d} A\}$
- $\mathcal{D}^+(A) = \{B : B \in Args, A \xrightarrow{d} B\}$
- $\mathcal{D}^\times(A) = \{B : B \in Args, A \xleftrightarrow{\times} B\}$

$\mathcal{D}^-(A)$ is the set of defeaters of A, $\mathcal{D}^+(A)$ is the set of arguments defeated by A, and $\mathcal{D}^\times(A)$ is the set of arguments in a controversial relation with A.

It is easy to see that the following properties hold:

- $\mathcal{D}^-(A) \subseteq U_\sigma(A)$,
- $\mathcal{D}^+(A) \subseteq D_\sigma(A)$ and
- $\mathcal{D}^\times(A) \subseteq I_\sigma(A) \cup E_\sigma(A)$

2.4 Semantics

The set of accepted arguments in the framework can be obtained in a similar way to [3]. In order to identify these arguments, we need some previous definitions.

Definition 2.8. *A set of arguments S is conflict-free if $\forall A, B : A \in S, B \in S$ then $(A, B) \notin \mathcal{C}$*

Definition 2.9. *An argument A is defended by a set of arguments S, if S is conflict-free and for every argument D such that D is in conflict with A and $\sigma(A, D) \neq \{A\}$ then exists an argument B in S such that $\sigma(D, B) \neq \{B\}$ and B is in conflict with D.*

Simply put, if all the arguments in S are accepted, then A should be accepted too. The restriction on σ ensures that the argument defense is not carried on by weakest arguments. In [3], several sets of extensions have been introduced. The skeptical semantic is defined by the following function, which characterizes the set of accepted arguments in the framework.

Definition 2.10. *The function F is defined as follows:*

- $\mathcal{F} : 2^{Args} \longrightarrow 2^{Args}$
- $\mathcal{F}(S) = \{A : A \text{ is defended by } S\}$

The set of acceptable arguments called S_{Acc} is obtained as the least fixpoint of the function $\mathcal{F}$. An argument not belonging to this set is said to be *rejected*.

Example 2.2. *Let* $AF = (Args, \mathcal{C}, \sigma)$ *an argumentation framework such that*
$Args = \{A, B, C, D\}$,
$\mathcal{C} = \{(A, B)(B, C)\}$
$\sigma(A, B) = \{A\}, \sigma(B, C) = \{B\}$
There are two defeat relations $A \xrightarrow{d} B$ *and* $B \xrightarrow{d} C$. *The arguments* A *and* C *are in* S_{Acc}, *and* B *is rejected.*

In the previous example, argument A is *defending* argument C. A controversial relation may lead to reject the arguments involved, and even more.

Example 2.3. *Let* $AF = (Args, \mathcal{C}, \sigma)$ *an argumentation framework such that*
$Args = \{A, B\}$,
$\mathcal{C} = \{(A, B)\}$
$\sigma(A, B) = \emptyset$
There is no defeat relation defined on AF. There is only one controversial relation $A \overset{\times}{\longleftrightarrow} B$. *Neither* A *or* B *can be accepted.*

Example 2.4. *Let* $AF = (Args, \mathcal{C}, \sigma)$ *an argumentation framework such that*
$Args = \{A, B, C\}$,
$\mathcal{C} = \{(A, B)(B, C)\}$
$\sigma(A, B) = \emptyset, \sigma(B, C) = \{B\}$
There is only one controversial relation $A \overset{\times}{\longleftrightarrow} B$. *There is also the defeat relation* $B \xrightarrow{d} C$. *However,* $S_{Acc} = \emptyset$ *and all the arguments are rejected.*

As seen in the previous examples, if an argument A is not involved in a controversial relation, but is defeated by a controversial argument, then A will probably be rejected. In some cases, controversial relations are not leading to rejections.

Example 2.5. *Let* $AF = (Args, \mathcal{C}, \sigma)$ *an argumentation framework such that*
$Args = \{A, B, C\}$,
$\mathcal{C} = \{(A, B)(B, C)\}$
$\sigma(A, B) = \emptyset, \sigma(B, C) = \{C\}$
There is only one defeat relation defined on AF: argument C *defeats argument* B. *There is only one controversial relation* $A \overset{\times}{\longleftrightarrow} B$. *However,* B *is defeated by* C, *so* B *can not be accepted. As* B *is involved in a controversial relation with* A, *and* B *is a rejected argument, then now* A *can be accepted.*

When a controversial argument is rejected by an accepted argument, then we say the controversial relation is *annulled*, because the status of the controversial arguments is finally determined.

Definition 2.11. *A controversial relation* $A \overset{\times}{\longleftrightarrow} B$ *is said to be annulled if there is an acceptable argument* C *such that* C *defeats* A *or* B.

An annulled controversial relation may cause other controversial relations to be annulled too. This is possible when two relations share a controversial argument.

Definition 2.12. *A controversial path from* A *to* B*, denoted* $A \overset{\times}{\rightleftharpoons} B$*, is a sequence of arguments* $A_1, .., A_n$ *such that* $A_1 = A$*,* $A_n = B$ *and* $A_i \overset{\times}{\longleftrightarrow} A_{i+1}, \forall\, i \in [1..n{-}1]$.

Due to the symmetric property of the controversial relation, if there is a controversial path from A to B, then there is also a controversial path from B to A.

Definition 2.13. *For any argument* A*, the controversial area of* A *is the set*

$$\varphi_\times(A) \;=\; \{B:\; A \overset{\times}{\rightleftharpoons} B\}$$

It is possible to define sets of arguments with equal controversial area. In fact, if B is in $\varphi_\times(A)$ then A is also in $\varphi_\times(B)$

Definition 2.14. *A controversial kernel of the argumentation framework* AF *is the set of arguments with equal* $\varphi_\times$.

Example 2.6. *The controversial kernel of the example 2.4 is* $\{A, B\}$*. The controversial kernel of example 2.5 is also* $\{A, B\}$

Example 2.7. *Let* $AF = (Args, \mathcal{C}, \sigma)$ *an argumentation framework such that*
$Args = \{A, B, C, D, E\}$,
$\mathcal{C} = \{(A, B)(B, C)(C, E)(D, A)\}$
$\sigma(A, B) = \sigma(B, C) = \emptyset, \sigma(E, C) = \{C\}, \sigma(D, A) = \{D\}$
In this framework, $A \overset{\times}{\longleftrightarrow} B$ *and* $B \overset{\times}{\longleftrightarrow} C$*. There is also the defeat relations* $C \overset{d}{\longrightarrow} E$ *and* $D \overset{d}{\longrightarrow} A$*. The only controversial kernel is* $\{A, B, C\}$

An argumentation framework may have more than one controversial kernel. These sets are formed only by controversial arguments. The following definition is a generalization of annulled controversial relations.

Definition 2.15. *A controversial kernel* K *is annulled if at least one argument in* K *is defeated by an acceptable argument.*

In other words, if a controversial relation $A \overset{\times}{\longleftrightarrow} B$ is annulled, then all the controversial relations in $\varphi_\times(A)$ are also annulled.

It is easy to see that a criterion is determinant in the outcome of the argumentation framework. In the next section we define the concept of *improvement* of a comparison criterion σ.

3 Improving the criterion

As stated before, a uniform comparison criterion is able to solve any conflict in the framework. No argument is rejected due to controversial situations. It is also possible to find a refinement of a given non-uniform criterion in order to minimize the set of rejected controversial arguments [6]. In this section, we explore de concept of improvement of σ and the impact of such refinement in the set of accepted arguments.

Definition 3.1. *An argument comparison criterion σ_1 is an improvement of the criterion σ_2 if the following properties hold*

1. *For any pair of arguments A and B, if $\|\sigma_2(A,B)\| = 1$ then $\sigma_1(A,B) = \sigma_2(A,B)$*
2. *There exists a pair of arguments C and D such that $\|\sigma_2(C,D)\| \neq 1$ and $\|\sigma_1(C,D)\| = 1$*

If σ_1 is an improvement of σ_2, then $I_{\sigma_1}(A) \cup E_{\sigma_1}(A) \subset I_{\sigma_2}(A) \cup E_{\sigma_2}(A)$ for some argument $A \in Args$.

Proposition 3.1. *If σ_1 is an improvement of a comparison criterion σ_2 then the set of accepted arguments under σ_2 is included in the set of accepted arguments of σ_1, under the skeptical semantic.*

An improved criterion probably translates a controversial situation in a defeat relation. Note that even when some pair of arguments A and B may not longer be incomparable, the outcome of the framework may change only if these arguments are in conflict. In other words, the improved criterion may decide over non conflictive pairs of arguments.

Definition 3.2. *Let $AF = (Args, \mathcal{C}, \sigma_1)$ an argumentation framework and σ_2 an improvement of the criterion σ_1. We say that σ_2 is a relevant improvement if there is a conflictive pair $(A,B) \in \mathcal{C}$ such that A and B are controversial arguments under σ_1 but not under σ_2.*

It is possible to define a new kind of arguments: those accepted when an improvement of the comparison criterion turns them accepted. This is called *conditional acceptance*. An argument A may be *conditionally accepted* if there is an improvement σ' of σ such that A is accepted under σ' but not under σ.

Definition 3.3. *An improvement σ_1 of a comparison criterion σ_2 is irrelevant if the set of arguments accepted under σ_1 is equal to the set of arguments accepted under σ_2.*

The set of conditionally accepted arguments in an improvement σ' is an extension to the arguments already accepted in the framework. In fact, conditionally accepted arguments are arguments in some preferred extension of [3].

Proposition 3.2. *If an argument A is conditionally accepted then A is in at least one preferred extension of AF, and not in the grounded extension.*

There is a relation between controversial situations and irrelevant comparison criterions, and it is based on a special situation called fallacy. Fallacies are very common in defeasible argumentation. Different kinds of fallacies and the impact of such situations in the set of accepted arguments are studied in [4, 5]. We present here a new definition of fallacy, based on controversial relations.

Definition 3.4. *A fallacy is a non-annulled controversial relation*

In other words, fallacies are dialectically unsolved controversial situations. Fallacious arguments (those involved in a fallacy) are not accepted. The most important problem with fallacies is that arguments defeated only by fallacious arguments can not be accepted[3].

The next definition relates irrelevant criterions with fallacies in the framework.

Definition 3.5. *If an improvement σ_1 of a comparison criterion σ_2 is irrelevant then it decides only over annulled controversial relations.*

This a very important concept. It shows that annulled controversial relations are not really a problem, because even if we adopt a decision on controversial arguments, the outcome of the system is the same. It makes sense to improve the criterion in order to solve controversial situations. Therefore, if the criterion is irrelevant, then it is not adopting preferences over fallacious arguments. In other words, if σ_1 turns some controversial situation r into a defeat relation r', and this relations is not adding new accepted arguments, then r is not a fallacy.

4 Conclusions

In argumentation systems, 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. The basis of this analysis is a special kind of relation, included in almost every abstract argumentation framework, called *attack* or *defeat relation* [1, 2, 3, 8]. The defeat relation is based on a preference made between conflictive arguments. Recently, some abstract frameworks are including preference orders [1, 2]. We defined an abstract argumentation framework that includes a binary symmetric relation representing argument conflicts, and a function used to evaluate conflicting arguments. In this framework it is clear how defeat relations are constructed, even when the arguments are treated abstractly. We also introduced the notion of *controversial* relation, which appears only when no preference is made over conflictive arguments.

We also presented a basic classification of comparison criterions and its impact on the set of accepted arguments. Finally, we defined the notion of *improvement of a comparison criterion* and its relation with the fallacies presented in the framework.

[3]Usually called indecision propagation [6]

Bibliografía

[1] L. Amgoud, C. Cayrol. *A reasoning model based on the production of acceptable arguments*, Proceedings of the 8^{th} International Workshop on Nonmonotonic Reasoning NMR'2000.

[2] Leila Amgoud, Claudette Cayrol. *On the Acceptability of Arguments in Preference-based Argumentation.* pp 1-7 Gregory F. Cooper, Serafín Moral (Eds.): UAI '98: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, July 24-26, 1998, University of Wisconsin Business School, Madison, Wisconsin, USA. Morgan Kaufmann 1998

[3] 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.

[4] Diego C. Martínez, Alejandro J. García, Guillermo R. Simari. *Significancia de las Falacias en los sistemas de argumentacion rebatible*, Congreso Argentino de Ciencias de la Computacion CACIC 99. Tandil, Octubre de 1999.

[5] Diego C. Martínez, *Alternativas para evitar falacias en la argumentación rebatible: un análisis comparativo*, Primer Workshop de Investigadores en Ciencias de la Computación. Universidad Nacional del San Juan. San Juan, Junio de 1999.

[6] Diego C. Martínez, *The Role of Argument Comparison in Dialectical Argumentation*, Workshop de Investigadores en Ciencias de la Computación. Universidad Nacional de la Plata. La Plata, Mayo de 2000.

[7] Alejandro J. García and Guillermo R. Simari. *Defeasible Logic Programming: An Argumentative Approach.* Theory and Practice of Logic Programming, 2002. Accepted for publication.

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

[9] Devid Poole. *On the comparison of theories: preferring the most specific explanation.* Proceedings of IJCAI-85, pp 144-147.