Una semántica de aceptabilidad para sistemas argumentativos abstractos con conflictos múltiples

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

Abstract

Algunos formalismos de argumentación rebatible, como el definido en [Dung93] o en [Vree97], mantienen ciertos elementos sin especificar, por lo que se los denomina *sistemas argumentativos abstractos*. En este trabajo se presenta un sistema argumentativo abstracto como el utilizado en [Dung93], pero con un nivel de detalle adicional. Aquí, la relación de derrota (o *attacks* tal cual definida por Dung) es desglosada en sus dos conceptos básicos: el conflicto entre argumentos, y el orden de preferencia entre los mismos. En función de estos elementos se analiza en este trabajo la noción de conflictos extendida, permitiendo que más de dos argumentos entren en conflicto, y su correspondiente proceso de evolución a la noción de derrota, junto con las consecuencias de la imposibilidad de decidir entre argumentos conflictivos. En este sistema abstracto las relaciones de derrota (las cuales pueden darse entre varios argumentos) existen únicamente como consecuencia de un proceso de evaluación y comparación *exitoso* entre los mismos. De no ser así, la semántica argumentativa aquí definida contempla los conflictos como núcleo de posibles extensiones.

`Palabras claves:` argumentación rebatible, aceptabilidad de argumentos, semánticas argumentativas.

1 Sistema abstracto de conflictos múltiples

Un sistema argumentativo plantea la existencia de argumentos que sustentan diversas proposiciones, denominadas conclusiones. Estos argumentos pueden entrar en conflicto cuando existe una contradicción entre elementos correspondientes a sus construcciones. La existencia de estos conflictos es un indicativo de que algunos argumentos no pueden ser aceptados como válidos simultáneamente. Por ende, se requiere un análisis un poco más detallado de los argumentos en juego para *resolver* cada uno de estos conflictos. Este examen está fundamentado en el hecho de que los argumentos sustentan sus conclusiones con diferente fuerza. La puesta en evidencia de esto último es tarea del criterio de comparación, el cual establece un orden entre los argumentos.

Los conflictos se resuelven comparando cada uno de los argumentos involucrados, derivando esta comparación en una *posible* relación de derrota entre los argumentos. En síntesis, establecer

*Becario de la Comisión de Investigaciones Científicas (CIC) de la Provincia de Buenos Aires

una relación de derrota entre argumentos conflictivos es simplemente establecer una jerarquía entre ellos, de acuerdo al orden de preferencia, lo cual puede expresarse de la siguiente manera:

$$derrota = conflictos + comparación$$

En nuestro trabajo no se especificará un lenguaje lógico ni la forma de construir argumentos, tratando los mismos con el máximo nivel de abstracción. Nuestro sistema argumentativo abstracto consistirá únicamente de un conjunto de argumentos, un conjunto denotando los conflictos entre los mismos y una relación de preferencia entre los argumentos.

Definición 1.1. *Un sistema argumentativo abstracto es una terna $\langle AR, \mathcal{C}, \mathcal{O} \rangle$ donde AR es un conjunto de argumentos, $\mathcal{C} \subseteq \mathcal{P}(AR)$ denota los conflictos entre argumentos, y $\mathcal{O} \subseteq AR \times AR$.*

Los argumentos serán denotados con letras mayúsculas. El conjunto $\mathcal{O}$ representa un orden parcial entre argumentos determinado por algún criterio de comparación en especial. Los conflictos del conjunto $\mathcal{C}$ pueden ser *múltiples*. El conjunto $\{A_1, A_2, ..., A_n\} \in \mathcal{C}$ indica que los argumentos $A_1, A_2, ..., A_n$ poseen un conflicto entre sí.

En general, en cada formalismo de argumentación rebatible, el concepto de conflicto está fuertemente relacionado con la noción de contradicción que utilice el lenguaje. Como en pos de la abstracción en este trabajo no se especifica ningún lenguaje en particular, permitimos en el sistema la presencia de conflictos en los que participe cualquier cantidad de argumentos. En [Vree97] se utiliza en el lenguaje un mecanismo para denotar contradicción que permite conflictos de este tipo.

Definición 1.2 (Conflictos múltiples y binarios). *Sea $\Lambda = \langle AR, \mathcal{C}, \mathcal{O} \rangle$ un sistema argumentativo abstracto, y sea $S \subseteq \mathcal{C}$ un conjunto de argumentos conflictivos. Se dice que S representa un* `conflicto binario`*, si $\|S\| = 2$. Si $\|S\| = 1$ entonces representa un conflicto unario, en caso contrario se dice que S representa un conflicto múltiple. Si todos los conflictos en Λ son binarios, se dice que Λ es un* `sistema argumentativo de conflictos binarios`*. De la misma manera, si existe al menos un conflicto múltiple, se dice que es un* `sistema argumentativo de conflictos múltiples`*.*

El orden parcial $\mathcal{O}$ establecido entre los argumentos puede surgir como consecuencia de algún método de comparación, como por ejemplo una función definida sobre AR, o determinado en forma directa de acuerdo a información externa al sistema sobre el dominio de construcción de los argumentos. Indica siempre un orden de *preferencia* entre los argumentos, y es la clave de la resolución de conflictos.

Por simplicidad, si es suficiente conocer que $(A, B) \in \mathcal{O}$ usualmente se indicará $A \leq B$. Si además $(B, A) \in \mathcal{O}$ se denotará $A \equiv B$. Si no existe un par (A,B) para dos argumentos A y B, entonces se indicará $A \bowtie B$ y se dirá que los argumentos son incomparables. Cuando sea preciso distinguirlo, si $A \leq B$ y $B \not\leq A$ entonces se denotará $A < B$.

Ejemplo 1.1. *La terna $\langle AR, \mathcal{C}, \mathcal{O} \rangle$ donde $AR = \{A, B, C, D\}$, $\mathcal{C} = \{\{A, B\}, \{B, C, D\}\}$, $\mathcal{O} = \{(A, B), (C, D)\}$ es un sistema argumentativo abstracto, de conflictos múltiples.*

Se pueden utilizar los símbolos $^-$ y $^+$ para denotar los subargumentos y superargumentos respectivamente: si A es un argumento, A^- es un subargumento y A^+ es un superargumento de A. Cuando sea necesario referirse a más de uno, se utilizarán subíndices. El conjunto $\mathcal{C}$ debe ser

completo, en cuanto a las relaciones conflictivas entre subargumentos. Esto es, si se incluye en el sistema el argumento A^-, entonces el superargumento A debe heredar los mismos conflictos para que el sistema refleje coherencia. Cabe destacar, sin embargo, que no necesariamente debe hacerse referencia a subargumentos en el sistema.

2 Resolución de los conflictos

El primer paso ante la presencia de conflictos es proceder a comparar los argumentos conflictivos para poder determinar una preferencia entre ellos. En este sistema los conflictos pueden darse entre más de dos argumentos, por lo que la determinación de la relación de derrota en función del criterio de preferencia es un poco más compleja que para los sistemas argumentativos binarios.

Definición 2.1. *Una relación de derrota es un par (S_1, S_2) de conjuntos de argumentos, que se denota $S_1 \xrightarrow{d} S_2$, donde S_1 es el conjunto derrotador y S_2 el conjunto derrotado.*

En particular, en los sistemas argumentativos de conflictos binarios los conjuntos derrotador y derrotado son unarios respectivamente. Se dice entonces que el argumento A derrota al argumento B, si $\{A, B\} \in \mathcal{C}$ y $B \leq A$ de acuerdo al criterio de comparación.

Ejemplo 2.1. *Sea $\langle AR, \mathcal{C}, \mathcal{O} \rangle$ un sistema argumentativo abstracto de conflictos binarios donde $AR = \{A, B, C\}$, $\mathcal{C} = \{\{A, B\}, \{B, C\}\}$ y $\mathcal{O} = \{(A, B), (B, C)\}$. Como el argumento A está en conflicto con el argumento B, y según el criterio de comparación B es preferible a A, entonces el argumento B derrota al argumento A. Una situación similar sucede con B y C. Se obtienen las siguientes relaciones de derrota:*

$$\{B\} \xrightarrow{d} \{A\} \quad ; \quad \{C\} \xrightarrow{d} \{B\}$$

La razón por la cual en la representación de la relación de derrota (por el símbolo $\xrightarrow{d}$) los argumentos van entre llaves, en notación conjuntista, es que la derrota puede darse entre conjuntos de argumentos, como consecuencia de conflictos múltiples, cuyo significado se verá a continuación. Cuando algún conjunto involucrado posea un sólo argumento se obviarán las llaves por simplicidad, y se escribirá únicamente el argumento correspondiente (por ejemplo, A $\xrightarrow{d}$ B).

2.1 Conflictos múltiples

Resolver un conflicto múltiple no es tan simple como resolver un conflicto binario. El orden preferencial entre los argumentos puede llevar a diferentes situaciones, de no tan evidente solución. En esta sección se verá como resolver conflictos múltiples entre argumentos.

Para ilustrar el proceso de elaboración de una relación de derrota, definimos los grafos de preferencia de los argumentos. Sea $S = \{A_1, A_2, ..., A_n\}$ un conjunto de argumentos en conflicto, y sea $\mathcal{G}_S$ el grafo dirigido determinado por los argumentos de S y por el conjunto $\mathcal{O}$ que denota las preferencias entre los argumentos de S. El grafo $\mathcal{G}_S$ indica las relaciones de fuerza entre los argumentos de S, de forma tal que existe un arco del argumento A al argumento B si $B \leq A$. El arco puede interpretarse como "*es tanto o más fuerte que*". En el grafo de preferencias, el sentido del arco indica siempre el argumento más débil.

Dados dos argumentos A_i y A_j pueden suceder:

- Si existe un camino del argumento A_i a A_j, se dice entonces que A_i *es tanto o más fuerte que* A_j.
- Si existe un camino de A_i a A_j, pero no existe un camino de A_j a A_i, se dice entonces que A_i *es más fuerte que* A_j.
- Si existe un camino de A_i a A_j y de A_j a A_i, entonces A_i *y* A_j *poseen la misma fuerza relativa.*
- Si no existe ningún camino entre ambos argumentos, entonces A_i *y* A_j *son incomparables.*

Si ante un conflicto entre argumentos $A_1, ...A_n$, existe un argumento $A_i (1 \leq i \leq n)$ tal que $A_i < A_j$ para todo $j \neq i$, (es decir, existe en $\mathcal{G}_S$ un camino desde cualquier argumento hasta A_i) entonces A_i es derrotado por el resto de los argumentos como conjunción. Esto es, A_i es derrotado por $\{A_1, ..., A_{i-1}, A_{i+1}, ..., A_n\}$, pues todos ellos lo señalan como el más débil del grupo. De esta forma se establece una preferencia entre todos, que deriva en la relación de derrota que denotaremos:

$$\{A_1, ..., A_{i-1}, A_{i+1}, ..., A_n\} \xrightarrow{d} \mathrm{A}_i$$

Esto significa que de estar los argumentos $\{A_1, ..., A_{i-1}, A_{i+1}, ..., A_n\}$ todos aceptados, no puede estarlo A_i. La base de esta relación de derrota es que el argumento A_i es más débil que el resto de los argumentos, y por lo tanto de menor prioridad frente a ellos: puede ser aceptado sólo si algún $A_j, 1 \leq j \leq n$ no es un argumento aceptado.

Puede suceder, sin embargo, que no exista un único argumento más débil que el resto, tal vez porque $\mathcal{G}_S$ puede estar formado por subgrafos completamente disjuntos. De ser así, en cada subgrafo están los argumentos relacionados entre sí de alguna forma, dependiendo de las propiedades estructurales que maneje el criterio de comparación. La misma situación puede darse si, a pesar de no existir subgrafos disjuntos, existen algunos argumentos que son únicamente destino de algún arco, pero no origen de uno de ellos.

De una manera u otra, esto implica que no existe un único argumento más débil en el conjunto. Sin embargo, se puede agrupar a los argumentos que son señalados como débiles por sus compañeros, caracterizados por el siguiente conjunto

$$\mathcal{D}(S) = \{A_i : \nexists A_j, A_j < A_i\}$$

En $\mathcal{D}(S)$ todos los argumentos son incomparables o de igual fuerza relativa, por lo que no puede establecerse una preferencia entre ellos. Son los argumentos más débiles del grupo.

Se puede definir entonces una relación de derrota entre los conjuntos $S - \mathcal{D}(S)$ y $\mathcal{D}(S)$, denotada

$$S - \mathcal{D}(S) \xrightarrow{d} \mathcal{D}(S)$$

Por lo tanto, de estar los argumentos de $S - \mathcal{D}(S)$ *todos* aceptados, entonces los argumentos de $\mathcal{D}(S)$ no pueden estar aceptados *todos* al mismo tiempo. Los argumentos del lado derecho de la relación de derrota poseen un conflicto entre sí, sólo cuando todos los argumentos de la izquierda son aceptados. Este conflicto "condicionado" a la aceptación previa de otros argumentos es

inevitablemente irresuelto, por la forma en que fue determinado el conjunto derrotado, es decir, agrupando conjuntos de baja fuerza relativa e incomparables.

Nótese que si en el conjunto conflictivo S todos los argumentos son incomparables, entonces $S - \mathcal{D}(S) = \emptyset$. En este caso denotamos la relación de la siguiente manera

$$\{\} \xrightarrow{d} S$$

Esto significa que los argumentos de S no pueden estar aceptados todos a la vez. Por la naturaleza múltiple del conflicto, y la falta de información adicional para resolverlo (en cuanto a la comparación) no se puede establecer una relación efectiva de derrota entre ellos. Este tipo de relaciones, en donde el conjunto derrotador es vacío, no es considerada realmente una *relación de derrota*, pues en realidad pone en evidencia la existencia de conflictos irresolutos en el sistema. Más adelante veremos cómo es tratado esto en la semántica del sistema.

Cuando un conjunto de argumentos S_1 derrote a un conjunto de argumentos S_2, se graficará de acuerdo a los diagramas de la figura 2, dependiendo de la cardinalidad de los conjuntos involucrados.

Cuando sea necesario, pueden distinguirse los argumentos integrantes de los conjuntos, en la forma habitual.

2.2 Conflictos y derrotas

La traducción de situaciones de conflicto a situaciones de derrota (principalmente recíprocas) cuando no se pudo establecer una preferencia entre los argumentos conflictivos es usualmente errónea, pues el significado de la unión de todas las relaciones de derrota que pueden armarse no es igual al significado de la noción de conflicto irresuelto. Por ejemplo, en los sistemas argumentativos de conflictos binarios, la presencia de argumentos conflictivos de igual fuerza conclusiva se traduce en relaciones de derrota recíprocas, aún cuando en realidad no existe una preferencia real entre ellos.

En el sistema deben convivir los dos conceptos, los conflictos y las derrotas, interpretado este último como una evolución del primero. Lo único que permite construir relaciones de derrota es el criterio de comparación, y se realiza únicamente cuando la comparación es concluyente. Cuando éste falla en tomar una decisión, debe prevalecer el concepto de conflicto entre argumentos, y la semántica del sistema debe ser coherente con el mismo.

Ejemplo 2.2. *Sea $\langle AR, \mathcal{C}, \mathcal{O} \rangle$ un sistema argumentativo donde $AR = \{P, Q, R, S\}$, $\mathcal{C} = \{\{P, Q, R, S\}\}$ y $\mathcal{O} = \{(Q, R), (R, P), (S, P)\}$. Existe un único conjunto de argumentos conflictivos $C = \{P, Q, R, S\}$, donde*

$$\mathcal{D}(C) = \{Q, S\}$$

por lo que se tiene que

$$\{P, R\} \xrightarrow{d} \{Q, S\}$$

Si bien los cuatro argumentos están en conflicto, de acuerdo al criterio de comparación se establece que de estar P y R aceptados, entonces Q y S no pueden estarlo al mismo tiempo. Esto significa que se pueden obtener diferentes combinaciones de argumentos aceptados:

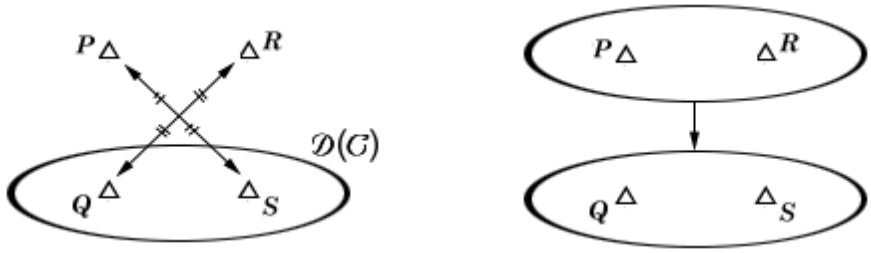

Figure 1: Conflictos y derrotas múltiples

$$\{P,R\},\ \{P,R,Q\},\ \{P,R,S\},\ \{P,Q,S\},\ \{R,Q,S\}$$

Esta situación corresponde al hecho de que existe una preferencia entre $\{P,R\}$ y $\{Q,S\}$, manifestada a través de una relación de derrota, mientras que Q y S están aún en conflicto, el cual es dependiente de P y R.

Cuando a partir de un conjunto de argumentos conflictivos S se identifican dos conjuntos S_1 y S_2 tales que $S_1 \xrightarrow{d} S_2$, como en el ejemplo anterior, puede decirse que los argumentos de S_2 están en *conflicto condicional* a la aceptación de los argumentos de S_1. En otras palabras, si se aceptan todos los miembros de S_1, se activa automáticamente un conflicto entre los miembros de S_2. Este conflicto, sin embargo, no puede ser solucionado debido a la falta de información para decidir entre argumentos de S_2, pues fueron precisamente seleccionados entre los más débiles de S.

Definición 2.2 (Clasificación de las relaciones de derrota). *Sean S_1 y S_2 dos conjuntos de argumentos tales que $S_1 \xrightarrow{d} S_2$. Decimos que la relación de derrota por ellos determinada es binaria, si $\|S_1\| = \|S_2\| = 1$, Terminante, si $\|S_1\| > 1$ y $\|S_2\| = 1$, Condicionante, si $\|S_1\| \geq 1$ y $\|S_2\| > 1$ o Nula, si $\|S_1\| = 0$*

Las relaciones de derrota binarias son las que se implementan prácticamente en todos los sistemas de argumentación rebatible [AG2k, AG97, BKT93, Sim92, PrakSart, PV98, Dung93]. Las relaciones de derrota terminantes prohíben la aceptación de un argumento concreto si fueron aceptados todos los argumentos derrotadores. Las relaciones condicionantes indican la existencia de un conflicto entre varios argumentos al aceptar un conjunto de argumentos disjunto. Las relaciones nulas son aquellas que denotan los conflictos irresolutos en el sistema. El significado de esta relación es el mismo que el de la noción básica de conflicto.

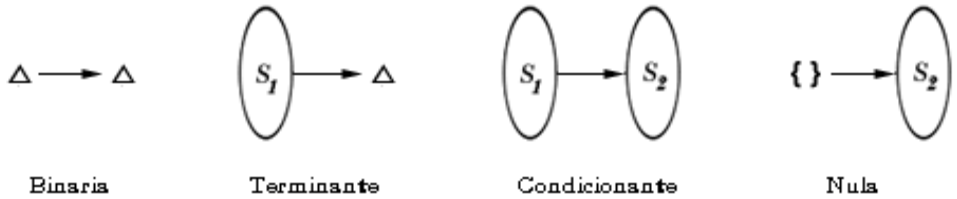

Figure 2: Diagramas de relaciones derrotas múltiples

La forma de graficar las relaciones de derrota múltiples pueden variar de acuerdo a su clasificación. Como se mencionó anteriormente, los conjuntos unarios pueden ser representados por el argumento que contienen, como en la figura 2, bajo la forma de un triángulo.

Ejemplo 2.3. *Sea* $\langle AR, \mathcal{C}, \mathcal{O} \rangle$ *un sistema argumentativo donde* $AR = \{A, B\}$, $\mathcal{C} = \{\{A, B\}\}$ *y* $\mathcal{O} = \{\}$. *Existe un único conjunto de argumentos conflictivos* $C = \{A, B\}$, *donde* $\mathcal{D}(C) = \{A, B\}$ *por lo que la derrota es la siguiente*

$$\{\} \xrightarrow{d} \{A, B\}$$

y por ende A *y* B *no pueden ser aceptados simultáneamente.*

En el conjunto de argumentos derrotados existen no sólo argumentos incomparables, sino también aquellos que poseen igual fuerza conclusiva.

Ejemplo 2.4. *Tomando el sistema argumentativo del ejemplo anterior, si*

$$\mathcal{O} = \{(A, B), (B, A)\}$$

entonces $\mathcal{D}(C) = \{A, B\}$ *y se obtiene el mismo resultado.*

Los conflictos no resueltos provocan inevitablemente la posibilidad de aceptar diferentes conjuntos de argumentos, sin poder decidir definitivamente entre ellos. Es por eso que el criterio de comparación tiene un fuerte impacto en la existencia de argumentos para los cuales es imposible tomar una decisión en cuanto a su aceptación o rechazo.

3 Semántica de aceptabilidad

La noción de aceptabilidad de argumentos es fundamental para la teoría de argumentación, pues surge del concepto básico de *defensa* entre argumentos. Se dice que un argumento A *defiende* a un argumento B, cuando el primero es derrotador de un derrotador del segundo. Por supuesto, B puede poseer más derrotadores, no necesariamente derrotados por A, por lo que A usualmente no es suficiente para defender en forma completa a B.

En un sistema argumentativo, muchos argumentos son defensores de otros argumentos. Es importante identificar estas *defensas*, puesto que aceptar los defensores de un argumento A implica sin duda aceptar al mismo argumento A.

En [Dung93], Dung describe informalmente el significado del concepto de argumentos aceptables, afirmando que "*para un agente racional G, un argumento A es aceptable si G puede defender a A de todos los ataques contra A*". Aquí presentaremos una noción análoga, adaptada a la existencia de conflictos múltiples entre argumentos.

Definición 3.1 (Aceptabilidad). *Un conjunto de argumentos Arg es aceptable con respecto a otro conjunto de argumentos S, si se cumple que:*

- *Para toda derrota* $S_1 \xrightarrow{d} S_2$, *si* $S_2 \subseteq S \cup Arg$, *entonces existe al menos un argumento* $d \in S_1$ *tal que existe un subconjunto* $B \subseteq S$ *tal que* $B \xrightarrow{d} d$.

Nótese que la noción de aceptabilidad está basada únicamente en las relaciones de derrota, pues los conflictos aún sin resolver no pueden ser determinantes en el sistema. Analicemos a continuación cuándo un argumento A es aceptable con respecto a un conjunto S, de acuerdo a las relaciones de derrota que pueden encontrarse.

En las derrotas múltiples, existe un punto en el cual el argumento A puede ser defendido: en al menos un argumento de cada conjunto derrotador. Es decir, en algún argumento de cada conjunto S_1 tal que

$$S_1 \xrightarrow{d} S_2$$

donde $A \in S_2$. Sin embargo, que esto falle no indica que A no pueda ser aceptado con respecto a S, pues existe otra situación en que los argumentos pueden *coexistir* y pertenecer a alguna extensión del sistema. La primer posibilidad de que A sea aceptado es que S derrote a sus derrotadores. Si A posee un conjunto como derrotador, entonces al menos un argumento de ese conjunto debe ser derrotado por S a través de una derrota *terminante*.

La segunda posibilidad es que, si S no puede defender a A, al menos sea compatible con S en el sentido de que no incluye a todos los argumentos que no pueden ser aceptados *junto* con A. Esta es la razón de la inclusión de la condición $S_2 \subseteq S \cup Arg$ en la definición 3.1. Dado que siempre $Arg \subseteq S_2$, esta condición no se cumpliría si existe en S un argumento que no pertenece a S_2.

En el caso de las relaciones de derrota binaria

$$\mathrm{D} \xrightarrow{d} \mathrm{A}$$

la única posibilidad de que se aplique la definición de aceptabilidad es que exista en S un derrotador de D, pues siempre se cumplirá que $S_2 \subseteq S \cup Arg$. Lo mismo sucede con las derrotas terminantes.

Observación 3.1. *A diferencia de Dung en [Dung93], la unión de conjuntos aceptables con respecto a un conjunto S no es un conjunto aceptable con respecto a S. Por ejemplo, en la derrota $A \xrightarrow{d} \{B, C, D\}$ el conjunto $\{B\}$ es aceptable con respecto a $\{A\}$. También lo es $\{C\}$ y $\{D\}$, pero no $\{B, C, D\}$.*

Ejemplo 3.1. *Sean A, B, C, D y E argumentos tales que las relaciones de derrota son las siguientes:*

$$A \xrightarrow{d} \{B, C, D\} \quad ; \quad E \xrightarrow{d} A$$

Sea $S_1 = \{B, C, D\}$. El argumento B es aceptable con respecto a los siguientes conjuntos:

- *El conjunto $\{E\}$, pues es E derrotador de A.*
- *El conjunto $\{C\}$, pues si bien éste no puede defender el ataque de A, el conjunto $\{C\} \cup \{B\} \not\supseteq S_1$.*
- *El conjunto $\{E, C, D\}$.*

No es aceptable con respecto al conjunto $\{C, D\}$ pues es equivalente a $S_1 - \{B\}$.

Informalmente la definción de *aceptabilidad* indica que en el único caso en que un argumento (en el ejemplo, B) puede ser aceptado con respecto a un conjunto S que contenga el resto de los argumentos que son derrotados en conjunto (C y D) es que además S contenga los derrotadores (E) de sus derrotadores (A).

Definición 3.2 (Conjunto admisible). *Un conjunto S de argumentos libre de conflicto se dice admisible si cada argumento en S es aceptable con respecto a S.*

Los conjuntos admisibles de argumentos sirven para varios propósitos. Primero, muestran qué es lo *necesariamente* justificado en un sistema argumentativo (aquellos argumentos que pertenecen a todos los conjuntos admisibles). Segundo, muestran qué está *posiblemente* justificado (aquellos que pertenecen al menos a un conjunto admisible). Y por último, muestran un escenario completo que satisface las restricciones de derrota y conflictos impuestas en el sistema. Las primeras dos opciones son usualmente denominadas posiciones escépticas y crédulas respectivamente.

Ejemplo 3.2. *Sea Λ un sistema argumentativo con los argumentos $\{A, B, C, D, E\}$ tal que posee las siguientes relaciones de derrota:*

$$C \xrightarrow{d} \{A, B\} \quad ; \quad E \xrightarrow{d} \{A, D\}$$

Los máximos conjuntos admisibles son $\{C, E, A\}$ y $\{C, E, B, D\}$.

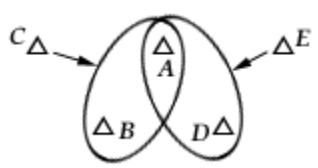

Figure 3: Conjuntos admisibles

Los argumentos E y C deben aceptarse pues al carecer de derrotadores son aceptables con respecto al conjunto vacío. El resto está en conflictos aún, por lo que no puede tomarse una decisión concreta. Si se decide aceptar A, no pueden aceptarse B y D, pues con ellos mantiene un conflicto al ser aceptado C y E. Si se decide rechazar A, entonces pueden aceptarse los argumentos B y D simultáneamente, pues estos son aceptables con respecto a $\{C, E\}$.

Los conjuntos admisibles indican los argumentos que *deben* aceptarse en el sistema junto con los argumentos conflictivos que *pueden* aceptarse. Al ser conflictos irresolutos, pueden existir diferentes combinaciones que extiendan el conjunto de argumentos aceptables.

Ejemplo 3.3. *Sea $\Lambda = \langle AR, \mathcal{C}, \mathcal{O} \rangle$ un sistema argumentativo tal que $AR = \{A, B, C\}$, $\mathcal{C} = \{\{A, B\}, \{A, C\}\}$ y $\mathcal{O} = \{(C, A)\}$. La relaciones de derrota de Λ son*

$$\{\} \xrightarrow{d} \{A, B\} \quad ; A \xrightarrow{d} C$$

Λ posee dos conjuntos admisibles maximales con respecto a la inclusión: $\{A\}$ y $\{B, C\}$.

Ejemplo 3.4. *Sea Λ un sistema argumentativo con las siguientes relaciones de derrota*

$$\{\} \xrightarrow{d} \{A, B\} \quad ; \quad \{A, B\} \xrightarrow{d} C$$

El conjunto $\{C\}$ es aceptable con respecto al conjunto vacío. La primer relación indica que A y B no serán aceptados simultáneamente, y precisamente eso es lo que se requiere para aceptar C. EL conjunto vacío cumple el papel de defensor de C. Los conjuntos $\{C, A\}$ y $\{C, B\}$ son conjuntos admisibles de argumentos.

No se obtiene el mismo resultado si se intercambia la segunda relación por $A \xrightarrow{d} C$, puesto que no existe un derrotador para A únicamente. El argumento A puede ser aceptado, y derrotar a C. El ejemplo 3.4 muestra como un conflicto irresoluto puede permitir la justificación de un argumento. En los sistemas argumentativos de conflictos binarios tradicionales, no se obtiene este resultado en base a la correspondiente semántica de aceptabilidad.

Ejemplo 3.5. *Sea Λ un sistema argumentativo con las siguientes relaciones de derrota*

$$\{\} \xrightarrow{d} \{A, B\} \quad ; \quad \{A, B\} \xrightarrow{d} \{C, D\} \quad ; \quad \{C, D\} \xrightarrow{d} \{E, F\}$$

La primer relación indica que los argumentos A y B no pueden ser aceptados simultáneamente. La segunda relación indica que si A y B son aceptados los dos a la vez, entonces no pueden ser aceptados C y D al mismo tiempo. Como esto nunca va a ocurrir, C y D pueden ser ambos aceptados. Por esa razón, y debido a la tercer relación de derrota, E y F no pueden ser aceptados al mismo tiempo. Por ende, los conjuntos $\{A, C, D, E\}$,$\{A, C, D, F\}$, $\{B, C, D, E\}$ y $\{B, C, D, F\}$ son conjuntos admisibles.

De los conjuntos admisibles, interesan principalmente aquellos que son maximales con respecto a la inclusión.

Definición 3.3 (Extensión preferida). *Una extensión preferida de $\langle AR, \mathcal{C}, \mathcal{O} \rangle$ es un conjunto maximal admisible de argumentos de AR.*

Ejemplo 3.6. *En el ejemplo 3.1, el máximo conjunto admisible es $\{E, B, C, D\}$. Si excluímos al argumento E del esquema, entonces existen varios conjuntos admisibles: $\{A, B, C\}$, $\{A, C, D\}$ y $\{A, C, B\}$. Debido a la elección hecha por el criterio de preferencia, A está en todos los conjuntos.*

3.1 Consecuencias del sistema argumentativo

Las extensiones preferidas son un indicativo de los argumentos que *podrían* eventualmente aceptarse en un sistema argumentativo Λ. Pero en principio es importante identificar los argumentos que *deben* ser aceptados en Λ.

La diferencia entre estos dos conjuntos de argumentos está en los argumentos en conflicto. Nada puede hacerse respecto a los conflictos irresolutos, por lo cual es natural no aceptar ninguno de los argumentos intervinientes. Aún así las posibles extensiones al sistema con estos argumentos son expresadas por medio de los conjuntos admisibles.

Se dice que un argumento A es *consecuencia* del sistema argumentativo, si el escenario de derrotas y conflictos entre argumentos justifica la aceptación de A, lo cual significa, dicho de otra manera, que A está justificado. Se puede definir un operador de punto fijo con el cual podamos

caracterizar el conjunto de argumentos que son *consecuencia* del sistema argumentativo en forma similar a [Dung93]. El principal problema aquí es que la unión de conjuntos de argumentos aceptables con respecto a un conjunto S no es necesariamente un conjunto aceptable con respecto a S.

Como se mencionó en la sección anterior, una forma de caracterizar los argumentos justificados es precisamente a través de los conjuntos admisibles:

Definición 3.4 (Argumentos justificados). *Sea Λ un sistema argumentativo. Los argumentos justificados de Λ son aquellos que pertenecen a la intersección de todos los conjuntos admisibles maximales.*

Pero también se puede obtener el conjunto de argumentos justificados de un sistema argumentativo por medio de una función monótona, en forma similar a [Dung93]. Para ello se requiere previamente un refinamiento del concepto de aceptabilidad de argumentos.

Definición 3.5 (Completamente aceptable). *Sea A un argumento y S un conjunto de argumentos. Se dice que A es* `completamente aceptable` *con respecto a S, si para toda relación de derrota $S_1 \xrightarrow{d} S_2$, tal que $A \in S_2$ se cumple que S_2 es aceptable con respecto a S.*

Ejemplo 3.7. *Sea Λ un sistema argumentativo con las siguiente relación de derrota:*

$$\{A\} \xrightarrow{d} \{B, C\}$$

El conjunto $\{B\}$ es aceptable con respecto a $\{A\}$. Lo mismo sucede con $\{C\}$. Sin embargo, no son completamente aceptables puesto que $\{B, C\}$ no es aceptable con respecto a $\{A\}$.

Definición 3.6 (Operador de consecuencia). *Para todo sistema argumentativo Λ se define la función F_Λ de la siguiente manera:*

$$F_\Lambda(S) = \{A : A \text{ es completamente aceptable con respecto a } S\}$$

Ejemplo 3.8. *En el ejemplo 3.7, $F_\Lambda(\emptyset) = \{A\}$, pues es el único completamente aceptable con respecto al conjunto vacío. De la misma forma, $F_\Lambda(\{A\}) = \{A\}$ pues B y C no son completamente aceptables.*

Proposición 3.1. *La función F_Λ es una función monótona.*

Demostración: F_Λ es una función monótona si para cualquier conjunto S_1, S_2 de AR, $F_\Lambda(S_1) \subseteq F_\Lambda(S_2)$ siempre que $S_1 \subseteq S_2$. Sea A un argumento de $F_\Lambda(S_1)$. Eso significa que A es completamente aceptable con respecto a S_1, es decir para toda regla $D_1 \xrightarrow{d} D_2$ donde $A \in D_2$ se cumple que existe una regla $X_1 \xrightarrow{d} X_2$ tal que $X_2 \subseteq D_1$ y $X_1 \subseteq S_1$. Pero entonces, si $X_1 \subseteq S_1$ se cumple también que $X_1 \subseteq S_2$. Por lo tanto A es defendido por S_2 en cada relación de derrota, lo que implica que $A \in F_\Lambda(S_2)$.

Definición 3.7 (Extensión básica). *Sea Λ un sistema argumentativo. La extensión básica de Λ, denotada G_Λ, es el menor punto fijo de la función F_Λ.*

Ejemplo 3.9. *En el ejemplo 3.5, $F_\Lambda(\emptyset) = \{C, D\}$ y $F_\Lambda(\{C, D\}) = \{C, D\}$, por lo que C y D son los argumentos justificados. Asi mismo, son los únicos que pertenecen a todas las extensiones preferidas.*

Ejemplo 3.10. *Sea Λ un sistema argumentativo con las siguientes relaciones de derrota*

$$\{\} \xrightarrow{d} \{A,B\} \quad ; \{A,B\} \xrightarrow{d} \{C\} \quad ; \{C\} \xrightarrow{d} \{D\} \quad ; \{D,E\} \xrightarrow{d} \{F\}$$

Con la aplicación reiterada de la función F_Λ se obtienen los siguientes conjuntos:

$$F_\Lambda(\emptyset) = \{C,E\} \quad ; \quad F_\Lambda(\{C,E\}) = \{C,E,F\} \quad ; \quad F_\Lambda(\{C,E,F\}) = \{C,E,F\}$$

por lo que los argumentos justificados son C, E y F.

4 Conclusiones

En este trabajo se presentó un sistema argumentativo abstracto en el cual, a diferencia de la mayoría de los sistemas, se permiten los conflictos entre múltiples argumentos. Estos conflictos *pueden* ser resueltos al igual que en los sistemas de conflictos binarios, evaluando los argumentos conflictivos de acuerdo a su fuerza conclusiva, conformando así una relación de derrota. Sin embargo, cuando esto no es posible, los conflictos permanecen como tales y conviven junto con las derrotas establecidas en el sistema. De esta forma se pueden capturar extensiones de argumentos que no existirían si los conflictos irresolutos se traducen a relaciones de derrotas recíprocas. De acuerdo a esto, se definió una semántica de admisibilidad extendida a conflictos múltiples, para capturar el conjunto de argumentos justificados en el sistema.

Bibliografía

[AG2k] A. J. García. Defeasible Logic Programming: Language, Operational Semantics and Parallelism PhD Thesis, Departamento de Ciencias de la Computación, Universidad Nacional del Sur, Bahía Blanca, Argentina. Diciembre de 2000.

[AG97] Alejandro García. *La programación en lógica rebatible. Su definición teórica y computacional* Tesis de Magister en Cs. de la Computación. Universidad Nacional del Sur. 1997.

[BKT93] A. Bondarenko, F. Toni, R.A. Kowalski. *An assumption-based framework for non-monotonic reasoning.* Proc. 2nd. International Workshop on Logic Programming and Non-monotonic Reasoning. MIT Press (1993), pág. 171-189.

[Dung93] P. M Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmotonic reasoning and Logic Programming and N-persons games, International Joint Conference on Artificial Intelligence 93 852-857, 1993.

[PrakSart] Henry Prakken y Giovanni Sartor. *Argument-based extended logic programming with defeasible priorities.* Journal of Applied Non Classical Logic 7, pág. 25-75.

[PV98] Henry Prakken y Gerard Vreeswijk. *Logics for Defeasible Argumentation.* Survey in defeasible argumentation. D. Gabbay (ed.), Handbook of Philosophical Logic, second edition. Kluwer Academic Publishers.

[Sim92] G. R. Simari y R. P. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53: 125–157,1992.

[Vree97] Gerard A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.