EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS XXII JORNADAS

VOLUMEN 18 (2012)

Luis Salvatico
Maximiliano Bozzoli
Luciana Pesenti

Editores

ÁREA LOGICO-EPISTEMOLÓGICA DE LA ESCUELA DE FILOSOFÍA
CENTRO DE INVESTIGACIONES DE LA FACULTAD DE FILOSOFÍA Y HUMANIDADES
UNIVERSIDAD NACIONAL DE CÓRDOBA

Epistemología e Historia de la Ciencia • Volumen 18 - 2012

Pruebas para la justificación de argumentos basadas en juegos

Gustavo Bodanza *

Introducción: la argumentación abstracta

La *argumentación abstracta* es una teoría de la argumentación rebatible en la que se analiza la interacción entre argumentos que se atacan unos a otros sin importar otros detalles. Un *marco argumentativo* (Dung, 1995) es modelado simplemente como un par $<A, R>$ donde A es un conjunto de entidades abstractas, los *argumentos*, y R es una relación binaria arbitraria sobre A, la relación de *ataque*. Mediante marcos argumentativos se pueden modelar situaciones como la siguiente:

a: *Cadena 3* dice que hoy va a llover en La Falda, luego se puede inferir tentativamente que lloverá en La Falda.

b: *La Voz del Interior* dice que hoy va a estar soleado en La Falda, luego se puede inferir tentativamente que no lloverá en La Falda.

c: *Cadena 3* suele ser más confiable que *La Voz del Interior*, luego se puede inferir tentativamente que *La Voz del Interior* está equivocada respecto del tiempo para hoy en La Falda

Si admitimos que los argumentos a y b se atacan mutuamente y que el argumento c ataca al argumento b, podemos representar esta situación mediante el marco argumentativo

$<\{a, b, c\}, \{(a, b), (b, a), (c, b)\}>$.

Sobre los marcos argumentativos se pueden definir distintos criterios para determinar cuáles serán los argumentos mejor justificados según los ataques que se dan en R, dando lugar así a distintas *semánticas de extensiones*. Es decir, cada semántica define una clase de conjuntos de argumentos, las *extensiones*, que representan los argumentos mejor justificados en el marco argumentativo.

Dung (1995) define una variedad de semánticas basadas en la noción de *aceptabilidad*: dado un marco argumentativo $<A, R>$, un argumento a ($a \in A$) es *aceptable* con respecto a un conjunto de argumentos S ($S \subseteq A$) sii $\forall b\ [(b \in A\ \&\ (b, a) \in R) \rightarrow \exists c\ (c \in S\ \&\ (c, b) \in R)]$ (i.e., S puede "defender" a a de cualquier ataque que reciba). A partir de aquí Dung define, entre otros, los siguientes tipos de *extensiones*:

preferida (preferred): cualquier conjunto máximamente (c.r.a $\subseteq$) admisible. Un conjunto S de argumentos es admisible sii todo argumento en S es aceptable c.r.a S y S está libre de conflictos, i.e. no hay dos argumentos a y b en S tales que $(a, b) \in$ R.

fundada (grounded): el menor conjunto S tal que $F(S) = S$, donde $F(S) = \{a$: a es aceptable c.r.a $S\}$.

completa (complete): cualquier conjunto S libre de conflictos tal que $F(S) = S$.

* U.N.Sur, CONICET, bodanza@gmail.com

Toda extensión preferida es completa, al igual que la fundada. La extensión fundada está incluida en toda extensión preferida -ver Dung, 1995.

Por ejemplo, dado el marco argumentativo <{*a*, *b*}, {(*a*, *b*), (*b*, *a*)}>, la semántica preferida produce las extensiones {*a*} y {*b*}, lo que puede interpretarse como que *a* y *b* están igualmente justificados y se puede elegir cualquiera de ellos indistintamente, mientras la semántica fundada produce la extensión vacía, lo que puede interpretarse como un rechazo de ambos argumentos. En este sentido, la semántica preferida representa un comportamiento "crédulo" y la fundada uno "escéptico". En algunos marcos argumentativos, sin embargo, ambas semánticas pueden coincidir. Por ejemplo, en <{*a*, *b*, *c*}, {(*a*, *b*), (*b*, *a*), (*c*, *b*)}>, tanto la semántica preferida como la fundada sancionan la única extensión {*c*, *a*}.

Propiedades de las semánticas de extensiones

Desde un enfoque más general, las distintas semánticas de extensiones se pueden clasificar según las propiedades que satisfacen. Baroni y Giacomin (2007) proponen, entre otras, las siguientes:

Admisibilidad Una semántica S satisface *admisibilidad* sii para cualquier marco argumentativo <*A*, R> y para toda extensión E de la semántica S, se cumple la condición:

$$a \in \mathrm{E} \rightarrow \forall b \in A\ [(b, a) \in R \rightarrow \exists c\ (c \in \mathrm{E}\ \&\ (c, b) \in R)]\ (1)$$

(i.e. todo argumento en una extensión es "defendible" (aceptable) con respecto a ella).

Admisibilidad Fuerte. Una semántica S satisface *admisibilidad fuerte* sii para cualquier marco argumentativo <*A*, R> y para toda extensión E de la semántica S, se cumple la condición:

$$a \in \mathrm{E} \rightarrow \forall b \in A\ [(b, a) \in R \rightarrow \exists c\ (c \in \mathrm{E}\ \&\ c \neq a\ \&\ (c, b) \in R)\ ]\ (2)$$

(i.e. todo argumento en una extensión es "defendible" con respecto a ella con un argumento distinto).

Restablecimiento. Una semántica S satisface *restablecimiento* sii para cualquier marco argumentativo <*A*, R> y para toda extensión E de la semántica S, se cumple la siguiente condición:

$$\forall b \in A\ [(b, a) \in R \rightarrow \exists c\ (c \in \mathrm{E}\ \&\ (c, b) \in R)] \rightarrow a \in \mathrm{E}\ (3)$$

(i.e., si *a* es defendible por la extensión entonces pertenece a la extensión).

Restablecimiento Débil. Una semántica S satisface *restablecimiento débil* sii para cualquier marco argumentativo <*A*, R> y para toda extensión E de la semántica S, se cumple la siguiente condición:

$$\forall b \in A\ [(b, a) \in R \rightarrow \exists c\ (c \in \mathrm{E}\ \&\ c \neq a\ \&\ (c, b) \in R)] \rightarrow a \in \mathrm{E}\ (4)$$

(i.e., si a es defendible por un argumento distinto de la extensión entonces pertenece a la extensión).

A estas propiedades vamos a agregar la siguiente:

No Conflictividad. Una semántica S es *no conflictiva* sii para cualquier marco argumentativo $<A, R>$ y para toda extensión E de la semántica S, se cumple la siguiente condición:

$$a,b \in \mathrm{E} \rightarrow (a, b) \notin R \quad (5)$$

Lema 1.
Si una semántica satisface admisibilidad fuerte entonces satisface admisibilidad.
Si una semántica satisface restablecimiento entonces satisface restablecimiento débil.

En lo que sigue introduciremos un juego argumentativo, llamado *juego de justificación*, que puede entenderse como una "teoría de pruebas" para la justificación de argumentos. La finalidad del trabajo será poner en correspondencia las nociones claves de este juego con las propiedades semánticas de Baroni & Giacomin.

Juegos de justificación

La caracterización estará basada en la elección estratégica de argumentos en un juego entre un proponente (P) y un oponente (O) tal que un argumento resulte justificado cuando, respetando las reglas del juego, P tenga una *estrategia ganadora* para defender a ese argumento. Primero definiremos un juego de justificación genérico, al cual se le podrán añadir nuevas reglas para capturar las distintas semánticas de extensiones.

Un *juego de justificación* sobre un marco argumentativo $<A, R>$ es un juego de suma-cero extensivo en el cual:

Hay dos jugadores, P y O, que van eligiendo acciones por turno.

La acción en cada nodo no terminal del árbol del juego consiste en la elección de un argumento $a \in A$.

Elecciones en los nodos de nivel par corresponden a P y aquellas en nodos impares corresponden a O.

El jugador P inicia el juego en el nodo de nivel 0 con un argumento que intentará defender de los ataques de O

La elección de un jugador i en un nivel $k > 0$ es un argumento a tal que existe un argumento b que es la elección del otro jugador $-i$ en el nivel $k - 1$ y $(a, b) \in R$.

El pago de P es 1 en los nodos terminales del árbol del juego si el nivel del nodo tiene numeración par, y en otro caso el pago es -1. El pago de O es 1 en los nodos terminales del árbol del juego si el nivel del nodo tiene numeración impar, y en otro caso el pago es -1

Las reglas que definen este juego pueden encontrarse en distintos trabajos, entre ellos, Modgil y Caminada (2008) y Viglizzo, Tohmé y Simari (2009), aunque éstos comprenden otras reglas adicionales que tienden a hacer más eficientes los juegos desde un punto de vista algorítmico. Aquí optamos por definir un juego con el mínimo de reglas posibles que permita establecer correspondencias con las propiedades de las semánticas de extensiones.

Una *historia* del juego es una secuencia $H = a_0, b^{H}_1, a^{H}_2, \ldots, a^{H}_{2k}, b^{H}_{2k+1}, \ldots$ (argumentos con subíndice par jugados por P, con subíndice impar jugados por O). Una *estrategia* para un jugador i es una función entre el conjunto de todos los nodos no terminales del árbol de todas las posibles jugadas y A, tal que a cada nodo correspondiente al otro jugador le asigna un argumento según las reglas del juego. El jugador i tiene una estrategia *ganadora* si para cualquier estrategia elegida por $-i$ todo juego finaliza en un nodo terminal que le asegura a i el pago 1.

Está claro que de todas las historias posibles que puede generar un juego, cada estrategia elegida por un jugador podría determinar una partición que separa las historias que no se jugarán de aquellas entre las cuales se encuentra la que efectivamente se jugará. En particular, si el proponente P tiene una estrategia ganadora para defender un argumento a, esta estrategia determinará que no se jugará ninguna historia en la que P no resulta ganador (siempre que P sea suficientemente sagaz).

Relaciones entre los juegos de justificación y las propiedades de las semánticas de extensiones

Admisibilidad

Para cualquier historia H generada por la estrategia ganadora de P (si la tiene), sea H^{P} el conjunto que reúne a todos los argumentos que jugará P en H. Un resultado sencillo es que P tiene una estrategia ganadora para cada argumento de ese conjunto. Otro resultado que será importante es que H^{P} es un conjunto de argumentos libre de conflictos.

Lema 2. Si P tiene una estrategia ganadora G entonces P tiene una estrategia ganadora para todo argumento $a \in H^{P}$, para cualquier historia H generada por G.

Proposición 1. Si P tiene una estrategia ganadora G entonces para cualquier historia H generada por G, el conjunto H^{P} satisface la condición (5).

Ahora podemos ver qué propiedades satisface una semántica de extensiones si hay una estrategia ganadora para los argumentos que pertenecen a las extensiones que sanciona.

Proposición 2. Si P tiene una estrategia ganadora G para defender un argumento a entonces para cualquier historia H generada por G, H^{P} satisface las condiciones (1) y (2).

Prueba:

(1): Supongamos que P tiene una estrategia ganadora G para a_0. Vamos a probar que H^{P} satisface la condición 1. Consideremos cualquier posible historia $H = a_0, b^{H}_1, a^{H}_2, \ldots, a^{H}_{2k}, b^{H}_{2k+1}, \ldots$ (argumentos con subíndice par jugados por P, con subíndice impar jugados por O) que pueda generar G y sea $H^{P} = \{a_0, a^{H}_2, \ldots, a^{H}_{2k}, \ldots\}$. Por hipótesis, existe un número natural n tal que la historia termina en a^{H}_{2n}. Este argumento es aceptable con respecto a H^{P} ya que no tiene atacantes. Esto implica también que a^{H}_{2n-2} es aceptable con respecto a $\{a^{H}_{2n}\}$, a^{H}_{2n-4} es aceptable con respecto a $\{a^{H}_{2n}, a^{H}_{2n-2}\}, \ldots$ y a_0 es aceptable con respecto a $\{a^{H}_{2n}, a^{H}_{2n-}$

$_2, \ldots, a^{H_2}\}$. O sea, todo argumento en H^P es aceptable con respecto a H^P; luego, H^P satisface la condición 1.

(2): Se puede ver que H^P satisface también la condición (2) teniendo en cuenta que ninguna historia puede tener ciclos de los cuales P no pueda escapar. Si este fuera el caso, entonces la historia no tendría un nodo terminal, ya que de ese modo O siempre tendría una jugada para la cual P no podría responder con un argumento no atacado. Pero esto contradice la hipótesis.

Por otra parte, que una semántica S satisfaga las condiciones (1) y (2) no implica que P tenga una estrategia ganadora para cualquier argumento de sus extensiones. Como contraejemplo podemos ver que en el marco argumentativo $<\{a, b, c\}, \{(a, c), (b, a), (c, b)\}>$, una semántica podría tomar como única extensión al conjunto $\{a, b, c\}$, que cumple la condición de admisibilidad fuerte, y sin embargo P no podría tener una estrategia ganadora para defender a ninguno de esos argumentos. Si en cambio la semántica cumple además con la condición de no conflictividad, la situación cambia:

Proposición 3. Si una semántica S satisface las condiciones (1), (2) y (5) entonces, si $a \in$ E para alguna extensión E de S entonces P tiene una estrategia ganadora para a.

Restablecimiento y función característica

Cualquier semántica cuyas extensiones sean puntos fijos de la función característica F de Dung (i.e. todas las completas, incluyendo la fundada y las preferidas) cumple la propiedad de restablecimiento (Baron & Giacomin, 2007). Si hay una estrategia ganadora para a entonces a pertenece a toda extensión de una semántica tal.

Lema 4. Si $S = F(S)$ entonces S cumple las condiciones (3) y (4).

Lema 5. Si P tiene una estrategia ganadora para a entonces $a \in S$ para todo $S \subseteq A$ tal que $S = F(S)$.

Prueba:

Sea $S \subseteq A$ tal que $S = F(S)$. Supongamos ahora que P tiene una estrategia ganadora G para a y sea H^P el conjunto de todos los argumentos jugados por P en una historia cualquiera H generada por G. Notemos que todos los nodos terminales que pertenecen a H^P son argumentos no atacados, luego todos ellos pertenecen a $F(\emptyset)$. Ahora bien, cualquier otro argumento $a' \in H^P$ es tal que para cualquier argumento b que ataca a a' existe un entero positivo i tal que algún argumento de $F^i(\emptyset)$ ataca a b, por lo cual $a' \in F^{i+1}(\emptyset)$. Por lo tanto, por construcción, a pertenece al menor punto fijo de F. En consecuencia, como la clase de puntos fijos de F forman un lattice[ii] con respecto a $\subseteq$, para todo conjunto $S \subseteq A$ tal que $S = F(S)$, $a \in$ E.

Proposición 4. Sea S una semántica tal que para toda extensión E se cumple $F(E) = E$. Si P tiene una estrategia ganadora para a entonces $a \in$ E para toda extensión E de S.

Prueba:
Se sigue inmediatamente de los lemas 4 y 5.

La conversa de esta proposición es falsa. Como contraejemplo podemos ver que el marco argumentativo <{*a*, *b*, *c*, *d*}, {(*a*, *b*), (*b*, *a*), (*a*, *c*), (*b*, *c*), (*c*, *d*)}> tiene dos extensiones preferidas: {*a*, *d*} y {*b*, *d*} (nótese que ambas son puntos fijos de *F*); sin embargo *d*, que pertenece a ambas, no puede ser defendido por una estrategia ganadora.

A partir de los resultados vistos, podemos obtener las siguientes conclusiones: De las proposiciones 1 y 2 se sigue que P tiene una estrategia ganadora *G* para defender un argumento *a* si y sólo si *a* pertenece a la extensión fundada y, por lo tanto, a toda extensión preferida; del lema 4 y la proposición 3 se sigue que si P tiene una estrategia ganadora para *a* entonces *a* pertenece a cualquier punto fijo de *F*.

Conclusión

Baroni & Giacomin (2007) han planteado una serie de propiedades o principios razonables con el fin de evaluar distintas semánticas de extensiones. En este trabajo hemos iniciado una contrapartida demostrativa de aquellas propiedades, basada en la construcción de juegos de justificación, que nos permitió poner en correpondencia las estrategias ganadoras del proponente con las propiedades semánticas de admisibilidad y restablecimiento.

Notas

i Aquí seguimos la convención en Teoría de Juegos de llamar a un jugador genérico *i* y al resto de los jugadores *-i*.

ii Ver Tarsky (1955) y cf. teorema 25 (3), Dung (1995: 330).

Bibliografía

BARONI, P.; GIACOMIN, M. On principle-based evaluation of extension-based argumentation semantics *Artificial Intelligence* **171** (10-15): 675-700, 2007.

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

MODGIL, S.; CAMINADA, M. Proof theories and algorithms for abstract argumentation frameworks, p. 105-132, en: RAHWAN, I. y SIMARI, G. (eds.), *Argumentation in AI*. Springer-Verlag, 2009.

TARSKI, A. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics* **5** (2): 285-309, 1955.

VIGLIZZO, I.; TOHMÉ, F.; SIMARI, G. The foundations of DeLP: defeating relations, games and truth values *Annals of Mathematics and Artificial Intelligence* **57** (2): 181-204, 2009.