

Dinámica en Argumentación Mediante Revisión de Creencias

Martín O. Moguillansky Nicolás D. Rotstein Marcelo A. Falappa
Alejandro J. García Guillermo R. Simari

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
Laboratorio de Investigación y Desarrollo en Inteligencia Artificial (LIDIA)
Departamento de Ciencias e Ingeniería de la Computación
Universidad Nacional del Sur (UNS)
Av. Alem 1253, (8000) Bahía Blanca, Argentina
e-mail: {mom, ndr, maf, ajg, grs}@cs.uns.edu.ar

1. Introducción

En este trabajo presentamos una línea de investigación que abarca la definición de una teoría abstracta que captura la dinámica de un marco argumentativo abstracto a través de la aplicación de conceptos de revisión de creencias. Esto comprende dos aportes novedosos: la definición de un marco argumentativo abstracto dinámico (DAAF, por *Dynamic Abstract Argumentation Framework*) y la aplicación de conceptos de revisión de creencias sobre el mismo para usufructuar su característica dinámica. El concepto de dinámica en el sistema aquí propuesto debe interpretarse como la fluctuación de argumentos dentro del DAAF: de acuerdo a ciertos eventos, algunos argumentos no son tomados en consideración para realizar inferencias. Por ejemplo, consideremos un sistema multi-agente basado en argumentación que controla la sincronización de semáforos. Si el sistema toma conocimiento de un corte de una avenida debido a reparaciones, no sólo formará argumentos para favorecer a las líneas de semáforos adyacentes a esta avenida, sino que no considerará ningún argumento para sincronizar los semáforos de la misma, ya que se encontrarán apagados.

Uno de los objetivos principales de esta línea es reificar los resultados obtenidos en formalismos argumentativos implementados. Esto resulta de gran importancia para tomar consideración de la aplicación práctica de dichos resultados. El formalismo elegido será Programación en Lógica Rebatible (DELP, por *Defeasible Logic Programming*) [4], debido a que se encuentra en un estado avanzado de desarrollo en nuestro laboratorio. Una investigación preliminar en la temática revisión/argumentación desde el punto de vista de su implementación en DELP fue publicada en [6] y una profundización de esta teoría aplicada sobre una primera versión del DAAF fue publicada posteriormente en [8]. Los resultados obtenidos en este último artículo fueron aplicados a DELP en un trabajo que se encuentra, actualmente, en revisión.

Es común encontrar en la literatura que los marcos argumentativos abstractos [7, 2] son construidos sobre el sistema propuesto por Dung [3], el cual no permite considerar dinámica. Es decir, se suele trabajar con marcos argumentativos que representan una “fotografía” del conocimiento que está siendo modelado. Para suplir esta falencia, presentamos el DAAF, el cual utiliza técnicas de revisión de creencias para describir la fluctuación de argumentos dentro del mismo; de alguna manera, dadas las “fotografías” inicial y final, proveemos la teoría para generar los “cuadros” intermedios. La teoría aquí presentada puede ser considerada abstracta desde dos puntos de vista: no sólo no hay restricción

a una representación lógica particular, sino que proveemos una caracterización de los operadores de cambio que no está circunscripta a implementación alguna.

2. Trabajo en Curso

Marco Argumentativo Abstracto Dinámico

La noción de argumento utilizada en esta línea de trabajo será absolutamente abstracta; es decir, los argumentos serán piezas de razonamiento que, a partir de un conjunto de premisas, llegan a una conclusión, sin referencia a una representación interna. Esto es importante para no generar complejidad innecesaria en la teoría.

A su vez, consideraremos una relación de *subargumento*, lo cual nos permitirá mantener la representación abstracta y, a su vez, dar a los argumentos cierta estructura mediante esta interrelación. En este sentido, el modelo gana naturalidad en la representación permitiendo que ciertas premisas sean logradas a través de otros argumentos (*i. e.*, subargumentos). En el DAAF, el conjunto de argumentos se considera *universal*, distinguiendo un subconjunto de él como *conjunto de argumentos activos*, a partir del cual el sistema realizará las inferencias. Finalmente, como en la noción usual de marco argumentativo, el DAAF cuenta con una relación de ataque entre argumentos para representar el concepto de derrota.

Definición 1 (Marco Argumentativo Abstracto Dinámico (DAAF)) *Un DAAF es una tupla $\Phi = \langle \mathbb{U}, \mathbb{A}, \mathbf{R}, \sqsubseteq \rangle$, donde \mathbb{U} es un conjunto finito de argumentos llamado **universal**, $\mathbb{A} \subseteq \mathbb{U}$ es llamado el **conjunto de argumentos activos**, $\mathbf{R} \subseteq \mathbb{U} \times \mathbb{U}$ denota una **relación de ataque** entre argumentos, y \sqsubseteq es un orden parcial sobre \mathbb{U} llamada **relación de subargumento**.*

Dada esta definición del DAAF, formalizaremos la noción de argumento.

Definición 2 (Argumento) *Un argumento es un conjunto de piezas de conocimiento interrelacionadas dando soporte a una conclusión a partir de evidencia y satisfaciendo: (**Consistencia Interna**) A posee consistencia interna respecto de \mathbf{R} ssi no existen $\mathcal{A}_i \sqsubseteq A$, $\mathcal{A}_j \sqsubseteq A$ tal que $\mathcal{A}_i \mathbf{R} \mathcal{A}_j$ o $\mathcal{A}_j \mathbf{R} \mathcal{A}_i$. (**Minimalidad**) A es minimal ssi A concluye α y no existe $\mathcal{A}_i \sqsubseteq A$ tal que \mathcal{A}_i concluye α .*

Cuando un argumento no cuenta con la evidencia suficiente para lograr soportar su conclusión, puede ser completado por otros argumentos que sí encuentren su propia evidencia. Estos argumentos que necesitan ser completados serán llamados *potenciales*, ya que su estado de activo o inactivo (de aquí en adelante nos referiremos a esto como *activación*) estará supeditado a la existencia de otros argumentos (activos) que logren concluir las premisas faltantes. El concepto de argumento potencial es independiente de su estado de activación.

La interacción entre argumentos y subargumentos con respecto a su activación se hará explícita al introducir el principio de **propagación de activación**, el cual le da coherencia a la dinámica del sistema:

(Propagación de Activación) $\mathcal{A} \in \mathbb{A}$ ssi para cada $\mathcal{A}_i \sqsubseteq \mathcal{A}$ se tiene que $\mathcal{A}_i \in \mathbb{A}$.

En palabras, un argumento está *activo* si y sólo si sus subargumentos lo están. A partir de los conjuntos \mathbb{U} y \mathbb{A} puede calcularse el conjunto \mathbb{I} de argumentos inactivos: $\mathbb{I} = \mathbb{U} \setminus \mathbb{A}$. De esta forma, podría razonarse acerca de conocimiento potencial.

La descripción del DAAF se completa indicando cómo se lleva a cabo el proceso para *garantizar* argumentos. En este caso, el mecanismo considerado se basa en *árboles de dialéctica* de la manera usual: se presenta un argumento de cuya conclusión se desea saber si es válido creer en ella;

este argumento va a ser la raíz del árbol de dialéctica. A continuación, se presentan todos aquellos *contraargumentos* para el argumento raíz, los cuales, a su vez, podrán ser derrotados por los *contraargumentos* correspondientes, y así siguiendo, hasta realizar un análisis exhaustivo. Cada rama del árbol de dialéctica será llamada *línea de argumentación*; sus elementos impares (como la raíz) serán llamados *argumentos de soporte*, mientras que los pares serán llamados *argumentos de interferencia*. Luego, un procedimiento de *marcado* determinará qué argumentos están derrotados y cuáles no (e. g., los argumentos hoja podrían estar marcados como no derrotados). La definición de la función de marcado determinará la semántica del marco, i. e., qué tipo de argumentos se consideran aceptados.

Ejemplo 1 El digrafo de argumentos en la Figura 1(a) ilustra un DAAF, donde los nodos triangulares son argumentos y los arcos denotan la relación de ataque, apuntando al argumento atacado. Se muestra el conjunto \mathbb{A} de argumentos activos, junto con el conjunto universal \mathbb{U} y el conjunto \mathbb{I} de argumentos inactivos, representados por triángulos de contorno punteado. Los subargumentos se ilustran como triángulos ubicados dentro de su respectivo superargumento. Nótese que el superargumento de A está inactivo debido a que tiene un subargumento inactivo.

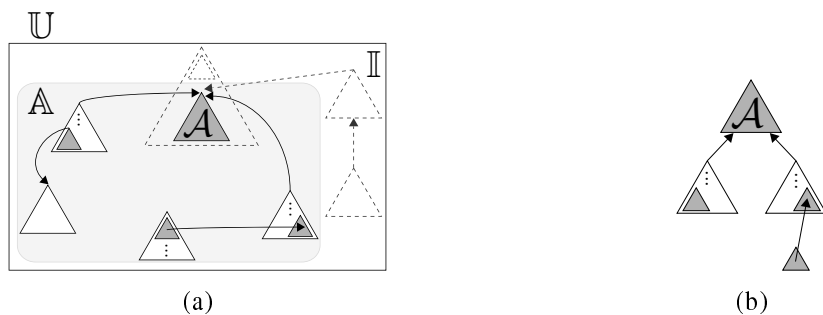


Figura 1: (a) Ejemplo de DAAF (b) Árbol de cubrimiento a partir de \mathcal{A}

La Figura 1(b) muestra el árbol de dialéctica de cubrimiento del grafo en (a), a partir del argumento \mathcal{A} . Obsérvese que, a pesar de que hay un ataque entre un argumento inactivo y un activo, los argumentos inactivos no serán considerados al analizar (i. e., marcar) el árbol. El marcado de este árbol de dialéctica nos permitirá determinar si el argumento raíz está garantizado. Por ejemplo, consideremos una función de marcado escéptica (m_{sk}) donde un dado nodo del árbol está no derrotado ssi es una hoja o sus derrotadores están derrotados. De acuerdo a m_{sk} , \mathcal{A} estaría derrotado. Este estado podría ser cambiado si se desactivara el derrotador de la izquierda de la raíz o ambos.

Teoría Argumentativa del Cambio

Respecto de la dinámica del marco argumentativo, resulta interesante la definición de operaciones de cambio reflejando las distintas situaciones de evolución posible. En particular, una operación necesaria es la activación de un dado argumento asegurando su garantía *a posteriori*. Esto es logrado mediante la denominada operación de revisión priorizada de argumento con garantía (WPA Revision “ \times^ω ”) que definiremos posteriormente.

En el contexto del árbol de dialéctica resulta necesario caracterizar aquellas líneas de argumentación (λ) responsables de la derrota del argumento raíz (\mathcal{A}). Estas serán denominadas *líneas de ataque* ($\text{Att}_{\mathcal{A}}$). En cada línea de ataque se selecciona (mediante la función “ γ ”) un argumento de interferencia a ser removido, de esta forma se logra garantizar al argumento raíz. La remoción de argumentos se lleva a cabo mediante una función de incisión “ σ ” que elige el conjunto de subargumentos apropiado a ser desactivado. Tanto las selecciones como las incisiones son guiadas por un principio de mínimo

cambio, tal como se especifica en la bibliografía clásica de teoría del cambio y revisión de creencias [1, 5].

Las incisiones sobre argumentos traen aparejada una dificultad: las *incisiones colaterales*. Una incisión colateral ocurre cuando la incisión efectuada sobre un argumento desactiva un subargumento que es parte de otro argumento en el árbol de dialéctica. Esto representa un problema, ya que amenaza la correctitud del proceso de revisión: una línea que no era de ataque puede convertirse en línea de ataque debido a una incisión colateral. Más aún, el argumento raíz podría resultar desactivado si no se toman los recaudos necesarios. Para subsanar esta problemática, se presentó la propiedad de *Preservación* [8] que determina el correcto funcionamiento de la función de incisión.

Definición 3 (Warrant-Prioritized Argument Revision) *Un operador de revisión WPA de un argumento $\mathcal{A} \in \mathbb{U}$ para un DAAF T es definido como $T \times^\omega \mathcal{A} = \langle \mathbb{U}, (\mathbb{A} \cup \{\mathcal{A}\}) \setminus \bigcup_i \sigma(\gamma(\lambda_i)), \mathbf{R}, \sqsubseteq \rangle$, donde $\lambda_i \in \text{Att}_{\mathcal{A}}$.*

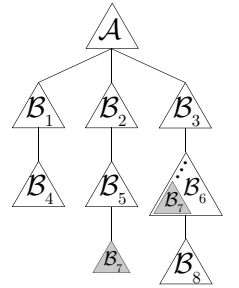
El operador de revisión WPA utiliza dos operadores para lograr efectuar la revisión: una expansión por un argumento $(\mathbb{A} \cup \{\mathcal{A}\})$ y una contracción “anti-garantía” $(\mathbb{A} \setminus \bigcup_i \sigma(\gamma(\lambda_i)))$, que elimina aquellos argumentos que obstaculizan la garantía de la raíz. Finalmente, el siguiente teorema constituye el principal resultado del trabajo en curso, cuya prueba puede encontrarse en [8].

Teorema 1

Si $T \times^\omega \mathcal{A}$ es la revisión del DAAF T por \mathcal{A} , entonces \mathcal{A} resulta garantizado.

Ejemplo 2

Consideremos una contracción anti-garantía realizada para garantizar \mathcal{A} sobre el árbol ilustrado en la figura de la derecha. En cuanto al criterio que guía a la selección, priorizaremos los argumentos inferiores, siguiendo un principio de mínimo cambio que intenta preservar la estructura del árbol. Por otra parte, utilizaremos la función de marcado m_{sk} y asumiremos que las líneas de ataque son aquellas cuya hoja es un argumento de interferencia. Haremos dos análisis: en el primero intentaremos no provocar incisiones colaterales, y en el segundo, las manejaremos apropiadamente, evitando que líneas que no son de ataque se transformen en líneas de ataque.



En cuanto a la línea $[\mathcal{A}, \mathcal{B}_1, \mathcal{B}_4]$ no hay incisión a ser realizada, ya que su hoja es un argumento de soporte. La línea $[\mathcal{A}, \mathcal{B}_2, \mathcal{B}_5, \mathcal{B}_7]$, por su parte, es de ataque y \mathcal{B}_2 debería ser seleccionado, dado que la desactivación de \mathcal{B}_7 provocaría una incisión colateral sobre \mathcal{B}_6 . Finalmente, en la línea $[\mathcal{A}, \mathcal{B}_3, \mathcal{B}_6, \mathcal{B}_8]$, \mathcal{B}_8 es seleccionado y desactivado. Las líneas de argumentación resultantes de la contracción son $[\mathcal{A}, \mathcal{B}_1, \mathcal{B}_4]$ y $[\mathcal{A}, \mathcal{B}_3, \mathcal{B}_6]$, de forma tal que \mathcal{A} resulta garantizado.

En el caso de permitir incisiones colaterales, la selección en la línea $[\mathcal{A}, \mathcal{B}_2, \mathcal{B}_5, \mathcal{B}_7]$ sería \mathcal{B}_7 , cuya desactivación afecta inevitablemente a \mathcal{B}_6 . La línea resultante luego de la incisión colateral sobre \mathcal{B}_6 es $[\mathcal{A}, \mathcal{B}_3]$, que ahora es una línea de ataque. Aplicando el principio de Preservación [8], el argumento \mathcal{B}_3 es seleccionado y desactivado. Finalmente, las líneas que resultan de la contracción son $[\mathcal{A}, \mathcal{B}_1, \mathcal{B}_4]$ y $[\mathcal{A}, \mathcal{B}_2, \mathcal{B}_5]$, por lo cual \mathcal{A} resulta garantizado.

3. Trabajo Futuro

Actualmente, estamos trabajando sobre la formalización del DAAF, la cual necesita ser refinada y completada con respecto a lo presentado en [8]. Otro artículo (actualmente en revisión) aplica los conceptos abstractos sobre el formalismo DELP, brindándonos el *feedback* necesario para mejorar la teoría definida hasta el momento.

Un análisis más profundo de los operadores de cambio es aún necesario, incluyendo la especificación de otros operadores, junto con la definición de un conjunto de postulados básicos y sus respectivos teoremas de representación. Otros operadores podrían ser:

- Expansión por una conclusión: activa todo argumento que posea dicha conclusión.
- Contracción por una conclusión: análogo a la expansión.
- Contracción por un argumento: desactiva un dado argumento.
- Revisión por expansión de derrotadores: activa los argumentos necesarios para garantizar el argumento raíz.

Una vertiente interesante de esta teoría estudiaría la variación de la relación de ataque para representar cambios en la preferencia: si un argumento \mathcal{A} actualmente derrota a \mathcal{B} , en el futuro esta relación podría revertirse, o podrían quedar bloqueados. Este cambio provocaría cambios en el conjunto de argumentos garantizados. Poder representar cambios en las preferencias de un agente o en las “reglas del juego” le brindaría gran flexibilidad a nuestra teoría. Por otra parte, poder efectuar cambios controlados (pero no arbitrarios) en las preferencias en dirección a una meta sería una herramienta interesante para modelar la evolución de un DAAF.

La aplicación de todos estos conceptos a sistemas que no fueron originalmente concebidos para actuar de acuerdo a la teoría argumentativa (como ser evolución de ontologías y arquitecturas de agentes) se encuentra actualmente bajo estudio. En paralelo con esta línea, se está investigando el contacto de este trabajo con el área de lógicas temporales para el análisis de la teoría desde una perspectiva global de su evolución: la noción de estados de evolución podría relacionarse a la elección de una dirección específica para las selecciones de argumentos e incisiones sobre ellos.

Referencias

- [1] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *The Journal of Symbolic Logic*, 50:510–530, 1985.
- [2] C. Chesñevar, A. Maguitman, and R. Loui. Logical Models of Argument. *ACM Computing Surveys*, 32(4):337–383, December 2000.
- [3] P. Dung. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning and Logic Programming and n -person Games. *Artificial Intelligence*, 77:321–357, 1995.
- [4] A. García and G. Simari. Defeasible logic programming: An argumentative approach. *Theory Practice of Logic Programming*, 4(1):95–138, 2004.
- [5] S. Hansson. *A Textbook of Belief Dynamics: Theory Change and Database Updating*. Springer, 1999.
- [6] M. Moguillansky, N. Rotstein, M. Falappa, and G. Simari. A Preliminary Investigation on a Revision-Based Approach to the Status of Warrant. In *Proc. of the 13th. Congreso Argentino de Ciencias de la Computación (CACIC 2007)*, pages 1536–1547, 2007.
- [7] H. Prakken and G. Vreeswijk. Logical Systems for Defeasible Argumentation. In D. Gabbay, editor, *Handbook of Philosophical Logic, 2nd ed.* Kluwer Academic Pub., 2000.
- [8] N. Rotstein, M. Moguillansky, M. Falappa, A. García, and G. Simari. Argument Theory Change: Revision Upon Warrant. In *Proc. of the 2nd. Conference on Computational Models of Argument (COMMA 2008)*, 2008.