EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS XIII JORNADAS

VOLUMEN 9 (2003), Nº9

Víctor Rodríguez
Luis Salvatico
Editores

ÁREA **L**OGICO-**E**PISTEMOLÓGICA DE LA **E**SCUELA DE **F**ILOSOFÍA
CENTRO DE **I**NVESTIGACIONES DE LA **F**ACULTAD DE **F**ILOSOFÍA Y **H**UMANIDADES
UNIVERSIDAD **N**ACIONAL DE **C**ÓRDOBA

Revisión de Creencias para Equipos de Agentes

Diego Letzen[*]

De los agentes racionales se espera que, obteniendo información sobre su entorno, incorporen esa información en teorías o conjuntos de creencias sobre esos dominios y los mantengan actualizados, revisando esas teorías en presencia de nueva información. El objetivo principal de la teoría AGM de dinámica de creencias ha sido modelizar este tipo de conducta.

De estos agentes, se espera asimismo que sean capaces de integrar grupos que compartan un marco general con el que abordar un área específica para resolver problemas.

El problema que se aborda en este trabajo es como representar situaciones de revisión de creencias en contextos de múltiples agentes, dentro del marco de dinámica de creencias conocido como modelo AGM.

La propuesta principal es un modelo de múltiples agentes integrados en una estructura general llamada *equipo* de agentes, definida por una relación sobre un conjunto de agentes, y que determina un conjunto de restricciones para las operaciones de estos agentes.

En los modelos o teorías de dinámica de creencias que participan de la tradición AGM, la representación de los estados epistémicos, o de creencias, de los agentes se realiza por medio de una función llamada de soporte por un conjunto de enunciados (aquellos que el agente admitiría) definidos sobre un lenguaje $\mathcal{L}$, cerrados bajo la operación de consecuencia clásica.

Se supone que son posibles tres actitudes epistémicas básicas: aceptación, rechazo e indeterminación, según si para un elemento α se tiene respectivamente $\alpha \in K$, $\sim\alpha \in K$ o, $\alpha \notin K$ y $\sim\alpha \notin K$; y en función de estas tres actitudes básicas, tres pares de tipos de cambio de creencia posibles: expansión (+) (de indeterminado a aceptado o de indeterminado a rechazado), revisión (*) (de aceptado a rechazado o de rechazado a aceptado) y contracción (–) (de aceptado a indeterminado o de rechazado a indeterminado).

Así como la mayoría de los cambios de creencias razonablemente puede verse como una sucesión de contracción expansión, formalmente, la operación de revisión puede definirse mediante lo que se conoce como identidad de Levi, con la siguiente expresión:

$$K * \alpha = (K - \sim\alpha) + \alpha$$

Es decir que, revisar un conjunto de creencias por un enunciado α, consiste en eliminar en primer lugar aquellos elementos del conjunto que pueden implicar $\sim\alpha$, para posteriormente expandir el conjunto resultante por α, con la certeza de no poner en peligro la consistencia del conjunto en cuestión.

[*] Universidad Nacional de Córdoba. CONICET

Las diferentes maneras de caracterizar los estados de creencias, cada una de estas operaciones, y los criterios de racionalidad que las determinan configuran los elementos de una teoría de cambio de creencias. Los elementos principales de esta configuración son los postulados que permiten caracterizar la dinámica en atención a ciertos principios o criterios de racionalidad. El aspecto dinámico de este modelo es expresado por estas operaciones de cambio y el modo en que estas funcionan esta representado por ciertas restricciones comprometidas con la racionalidad del cambio propuesto.

La preservación de la consistencia se torna una guía de la dinámica de creencias porque, por definición, un conjunto de creencias está cerrado bajo consecuencia lógica, $K = Cn(K)$ donde Cn es la operación de cierre lógico. Esta operación no es otra que la función de consecuencia que Tarski caracterizara en sus trabajos de la década del '30.

En la teoría AGM de cambio de creencias la noción de consecuencia Cn produce más resultados de los deseados. La operación de expansión, por ejemplo, es prácticamente inútil al pretender representar las incorporaciones, pues puede definirse simplemente en función de la operación de unión de conjuntos

$$K + \alpha = Cn(K \cup \{\alpha\})$$

Cargando con una gran cantidad de problemas de esa manera, por lo que se hace necesario complementarla o cubrirla por la de revisión o tal vez debiéramos llamarle de expansión consistente puesto que es la única que nos garantiza poder permanecer tranquilos al incorporar un elemento a nuestro conjunto de creencias sin que se trivialice.

Esta noción (Cn) queda caracterizada en términos generales por satisfacer

1. Inclusión: $A \subseteq Cn(A)$.
2. Monotonía: Si $A \subseteq B$, entonces $Cn(A) \subseteq Cn(B)$.
3. Iteración: $Cn(A) = Cn(Cn(A))$.

Y las siguientes tres propiedades:

- Si α puede ser deducida de A por medio de una instancia de tautología, entonces $\alpha \in Cn(A)$. (Supraclasicalidad).
- $\beta \in Cn(A \cup \{\alpha\})$ si y sólo si $(\alpha \rightarrow \beta) \in Cn(A)$. (Deducción).
- Si $\alpha \in Cn(A)$, entonces $\alpha \in Cn(A')$ para algún conjunto finito $A' \subseteq A$. (Compacidad).

De modo que la operación utilizada incluye a la noción veritativo funcional de consecuencia.

Como es usual en este contexto, consideramos a un conjunto A consistente si para ninguna α se tiene que $\{\alpha \wedge \sim\alpha\} \in Cn(A)$.

Se supone que de esta forma se tiene un modelo definido en base a ciertas pretensiones de racionalidad que permitirían representar la dinámica de creencias en agentes racionales.

La aplicación supuesta en forma estándar para este modelo son agentes individuales, que reciben estímulos del medio bajo la forma de entradas o 'inputs', y devuelven como resultado una nueva teoría o conjunto de creencias, típicamente consistente, que incorpora con el mínimo de variaciones, los nuevos elementos.

Una aplicación distinta y muy interesante es la relacionada con áreas nuevas como las llamadas inteligencias sociales, en alusión a contextos de interacción de agentes para la comprensión o solución de un problema.

Los dos problema principales relacionados con este nuevo enfoque son los de la consistencia, y el de la definición del mecanismo de revisión en estos contextos especiales. La inexistencia de normas incompatibles es un ideal racional análogo al de la consistencia de los enunciados que representan las creencias de un agente o de los elementos de una base de datos. Esto está reflejado en la utilización de la operación clásica de consecuencia para caracterizar los modelos de cambio.[1]

Una visión atenta de este aspecto del problema nos permite argumentar que muchos de los casos problemáticos, tal vez los más interesantes, se refieren a la cohabitación de partes parcialmente incompatibles en un todo, como sucede por ejemplo con la información proveniente de múltiples fuentes, o el sistema que resulta de la combinación de diversas leyes o códigos en la mente de un legislador ideal.

El principal aporte de este trabajo es la redefinición de las operaciones tradicionales de cambio para una estructura que resulta de la combinación de agentes a la que llamamos un equipo.

Un equipo es la forma de caracterizar un conjunto de agentes vinculados por una relación (intuitivamente la de integrar equipo con) la que en este caso debe ser una relación de equivalencia, que permite dirigir el resultado de las operaciones de cambio de creencias en forma conjunta.

Definimos para esto un subconjunto de operaciones de cambio restringidas por una relación definida sobre bases de creencias (agentes). Este conjunto de operaciones, limita los resultados posibles de las operaciones y permite dirigir el proceso de cambio para un conjunto de agentes, reduciendo con esto la amplitud de resultados para las operaciones de cambio, la incidencia de los mecanismos de selección y consiguientemente la complejidad de la determinación de los cambios de creencias.

Este trabajo se inscribe en un contexto más amplio en el que se pretende incorporar relaciones más complejas (que no sean de equivalencia) las que puedan caracterizar equipos variables y sobre todo jerarquizados. Se pretende además utilizar esta estructura para definir operaciones de alcance global, de revisión de creencias en equipos de agentes, modificando los elementos de un conjunto de agentes (los integrantes de un equipo), lo que acarrearía una dificultad extra, por la diferencia de resultados que la misma revisión puede acarrear para cada uno de los integrantes del equipo.

Se logra con esta modificación, no sólo dar cuenta de un proceso de cambio de creencias más complejo que el original; también se permite utilizar una estructura de representación de agencias más compleja (el equipo) que puede interpretarse a su vez como un agente y que permite representar situaciones de creencias inconsistentes.

A fin de completar esta presentación, consideremos estos conjuntos que representan dos bases de creencias (de dos potenciales integrantes de un equipo):

$K_1=\{\alpha, \alpha \rightarrow \beta, \gamma, \sim\beta \rightarrow \sim\gamma, \beta, \varepsilon\}$

$K_2=\{(\alpha \vee \delta) \rightarrow \beta, \delta, \alpha \vee (\beta \vee \sim\gamma), \sim\alpha, \gamma, \varepsilon \rightarrow \beta\}$

Una de las formas de realizar una contracción de una base de creencias es mediante la obtención de los mínimos conjuntos en función de la inclusión, que implican el elemento a contraer. Supongamos que deseamos contraer β. Los β-kernel de cada uno de ellos son:

$K_1 \Downarrow \beta = \{\{\alpha, \ \alpha \rightarrow \beta\}, \{\gamma, \sim\beta \rightarrow \sim\gamma\}, \{\beta\}\}$

$K_2 \Downarrow \beta = \{\{(\alpha \vee \delta) \rightarrow \beta, \delta\}, \{\alpha \vee (\beta \vee \sim\gamma), \sim\alpha, \gamma\}$

Una vez determinados los β-kernel, debemos aplicar una función, llamada de incisión, sobre sus elementos, a fin de "recortar" un elemento de cada uno de estos conjuntos implicadores e inhibir así la inferencia del elemento a contraer.

En el caso que nos ocupa, esto da lugar a las siguientes opciones de selección σ_1 para K_1 y σ_2 para K_2:

$\sigma_1 (K_1 \Downarrow \beta) = \{\alpha, \beta, \gamma\}$; $\sigma_1' (K_1 \Downarrow \beta) = \{\alpha, \sim\beta \rightarrow \sim\gamma, \beta\}$;

$\sigma_1'' (K_1 \Downarrow \beta) = \{\alpha \rightarrow \beta, \beta, \gamma\}$; $\sigma_1''' (K_1 \Downarrow \beta) = \{\alpha \rightarrow \beta, \sim\beta \rightarrow \sim\gamma, \beta\}$.

$\sigma_2 (K_2 \Downarrow \beta) = \{(\alpha \vee \delta) \rightarrow\beta, \alpha \vee (\beta \vee \sim\gamma)\}$, $\sigma_2' (K_2 \Downarrow \beta) = \{\delta, \sim\alpha\}$;

$\sigma_2'' (K_2 \Downarrow \beta) = \{(\alpha \vee \delta) \rightarrow\beta, \gamma\}$; $\sigma_2''' (K_2 \Downarrow \beta) = \{\delta, \gamma\}$;

$\sigma_2'''' (K_2 \Downarrow \beta) = \{(\delta, \alpha \vee (\beta \vee \sim\gamma)\}$; $\sigma_2''''' (K_2 \Downarrow \beta) = \{(\alpha \vee \delta) \rightarrow\beta, \sim\alpha\}$.

En total se debe elegir sobre 10 conjuntos, entre $4 \times 6 = 24$ configuraciones posibles.

Consideremos ahora en vez de esto que sería el abordaje más o menos ortodoxo, la incorporación de los siguientes elementos.

Sea R una relación, seguramente de equivalencia, definida en el conjunto $\mathcal{P}(\mathcal{L})$ cuyo significado intuitivo es "*formar equipo con.*"

Lo que esta relación hace es devolver un conjunto, el de los miembros del equipo R de agentes.

Definimos ahora sobre esta base, el conjunto $E \Downarrow \beta$, de los β-kernel de *E*:

Dado un equipo $E = \{K_1, K_2, \ldots, K_n\}$ y una fórmula β, es posible definir un conjunto $E \Downarrow \beta$ tal que $X \in E \Downarrow \beta$ sii

(1) $X \subseteq \bigcup_{i=1}^{n} K_i$

(2) $\beta \in Cn(X)$

(3) Si $Y \subset X, \beta \notin Cn(Y)$

Definimos ahora, para una relación R, una nueva función de incisión σ_R en un equipo $E = \{K_1, K_2, ..., K_n\}$ generado a partir de los elementos de R, tal que para cualquier β,

i)	$\sigma_R(E \Downarrow \beta) \subseteq \bigcup(E \Downarrow \beta)$
ii)	Si $\varnothing \neq X \in (E \Downarrow \beta)$, $X \cap \sigma_R(E \Downarrow \beta) \neq \varnothing$
iii)	$\sigma_R E$ es el menor conjunto respecto de la relación de inclusión que cumple las condiciones precedentes.

Por último, en el caso de realizarse una contracción,

Para todo $K \in E$, $K -_R \beta = K / \sigma_R(E \Downarrow \beta)$

Lo que esta función hace es 'cortar' un elemento de cada conjunto que implica β en cada una de las bases que integran el equipo, pero esta acción está dirigida mediante la cláusula iii) de forma que se restringe el campo sobre el que realizar la selección a los conjuntos con menor cantidad de elementos. O lo que es igual, la selección está dirigida para maximizar el acuerdo respecto de lo que se debe eliminar, minimizando las diferencias entre lo que cada agente puede hacer.

Veamos nuevamente el funcionamiento de la operación con un ejemplo concreto:
En el ejemplo que habíamos considero inicialmente, contábamos con 24 configuraciones distintas para resolver la contracción de β de las dos bases que conforman un equipo E.

$\boxed{\sigma_1(K_1 \Downarrow \beta) = \{\alpha, \beta, \gamma\}}$; $\sigma_1'(K_1 \Downarrow \beta) = \{\alpha, \sim\beta \rightarrow \sim\gamma, \beta\}$;

$\sigma_1''(K_1 \Downarrow \beta) = \{\alpha \rightarrow \beta, \beta, \gamma\}$; $\sigma_1'''(K_1 \Downarrow \beta) = \{\alpha \rightarrow \beta, \sim\beta \rightarrow \sim\gamma, \beta\}$.

$\sigma_2(K_2 \Downarrow \beta) = \{(\alpha \vee \delta) \rightarrow \beta, \alpha \vee (\beta \vee \sim\gamma)\}$; $\boxed{\sigma_2'(K_2 \Downarrow \beta) = \{\delta, \sim\alpha\}}$;

$\sigma_2''(K_2 \Downarrow \beta) = \{(\alpha \vee \delta) \rightarrow \beta, \gamma\}$; $\sigma_2'''(K_2 \Downarrow \beta) = \{\delta, \gamma\}$;

$\sigma_2''''(K_2 \Downarrow \beta) = \{(\delta, \alpha \vee (\beta \vee \sim\gamma)\}$; $\sigma_2'''''(K_2 \Downarrow \beta) = \{(\alpha \vee \delta) \rightarrow \beta, \sim\alpha\}$.

¿Qué deben priorizar los miembros de un equipo?
Si consideramos un mecanismo de selección global, el procedimiento se simplifica de la siguiente forma:

$(E \Downarrow \beta) \supseteq (K_1 \Downarrow \beta) \cup (K_2 \Downarrow \beta)$

$(E \Downarrow \beta) = \{\{\alpha, \alpha \rightarrow \beta\}, \{\gamma, \sim\beta \rightarrow \sim\gamma\}, \{\beta\}, \{(\alpha \vee \delta) \rightarrow \beta, \delta\}, \{\alpha \vee (\beta \vee \sim\gamma), \sim\alpha, \gamma\}, \{\varepsilon, \varepsilon \rightarrow \beta\}\}$

$$\sigma_R(\text{‘}E \Downarrow \beta) = \{\beta, \gamma\} \cup \left\{ \begin{array}{l} \{\alpha, \delta\} \\ \{\alpha \to \beta, \delta\} \\ \{\alpha, (\alpha \vee \delta) \to \beta\} \\ \{\alpha \to \beta, (\alpha \vee \delta) \to \beta\} \end{array} \right. \cup \left\{ \begin{array}{l} \text{uno de:} \\ -\varepsilon \\ -\varepsilon \to \beta \end{array} \right.$$

La opción de eliminar $\sim\alpha$ desaparece pues este elemento aparece ligado a γ, que es un candidato firme al igual que β a ser eliminado.

Finalmente, resta calcular $K_1 / \sigma_R(E \Downarrow \beta)$ y $K_2 / \sigma_R(E \Downarrow \beta)$

Suponiendo que $\sigma_R(E \Downarrow \beta) = \{\beta, \gamma, \alpha \to \beta, (\alpha \vee \delta) \to \beta, \varepsilon\}$

$K_1 -_R \beta = \{\alpha, \sim\beta \to \sim\gamma\}$

$K_2 -_R \beta = \{\delta, \alpha \vee (\beta \vee \sim\gamma), \sim\alpha, \varepsilon \to \beta\}$

Otra ventaja de esta forma de proceder es que elimina la posibilidad de que el elemento a contraer se siga infiriendo del conjunto de creencias del equipo en la situación en que yo tengo el mismo β-kernel en dos agentes, si puedo seleccionar un elemento distinto de cada uno de ellos, β seguiría infiriéndose del equipo en forma ‘emergente’.

Nota

1 El cierre bajo consecuencia clásica (*Cn*) tiene un papel importante no sólo en la definición de los estados, sino también en la de las operaciones puesto que estas siempre se hacen, no importa cual sea la presentación elegida (postulados, funciones de selección, semánticas...) sobre el horizonte de esta operación.

Bibliografía

Alchourrón, C.; Gärdenfors, P., y Makinson, D (1985), “On the Logic of Theory Change: Partial Meet Contraction and Revision Functions”, *Journal of Symbolic Logic*, 50, 510-530.

Gabbay, D.; y Woods, J. (2001), “The new logic”, *LJ of IGPL*, 9, 2.

Gärdenfors, P. (1988), *Knowledge in Flux. Modeling the Dynamics of Epistemic States*. Cambridge, Mass.: Bradford/MIT.

Malheiro, B.; Jennings, N.; y Oliveira, E. (1994), “Belief Revision in Multi-Agent Systems”, en *Proceedings of the 11th European Conference on Artificial Intelligence (ECAI'94)*, Amsterdam, Holland, Agosto 1994

Priest, G.; y Routley, R. (1989), “Systems of Paraconsistent Logic”, en Priest, G.; Sylvan, R.; y Norman, J. (eds.), *Paraconsistent Logic. Essays on the Inconsistent*. Munich: Philosophia Verlag.