EPISTEMOLOGÍA E HISTORIA DE LA CIENCIA

SELECCIÓN DE TRABAJOS DE LAS VII JORNADAS
1997

Patricia Morey
José Ahumada
Editores

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

Silvia B. Lerner
Facultad Cs.Ecs./U.B.A.

EXPLICABILIDAD Y RAZONAMIENTO REVOCABLE

1. Introducción

La búsqueda de reconstrucciones formales del razonamiento revocable o default ha conducido , por un lado, a sistemas que recurren a cambios de lógica (extendiendo la lógica mediante operadores y reglas o procedimientos no-monótonos) y, por otro, a sistemas como el de Poole, que reniegan de la necesidad de nuevas lógicas y reivindican la suficiencia de la lógica clásica. Poole sostiene que un caso característico de razonamiento default es el razonamiento hipotético (explicación a partir de hipótesis) de las teorías científicas.

En la presente comunicación se analizan propiedades formales de la relación de explicabilidad definida en la propuesta de Poole. Dicho análisis revelará que su sistema resulta más adecuado comparativamente que otros sistemas que persiguen objetivos similares.

2. Breve caracterización del sistema de Poole.

En su artículo de 1988 'A Lógical Framework for Default Reasoning', D. Poole presenta una propuesta para reconstruir el razonamiento revocable que se diferencia de propuestas similares en no apelar a cambios de lógica. La lógica es la clásica de primer orden, pués ' ... antes que ser un problema con la lógica, la no-monotonía es un problema acerca de cómo la lógica se usa'.[1]

El objetivo de Poole es construir un formalismo que refleje la explicación de hechos a partir de hipótesis en el marco de una teoría. Poole ve al razonamiento hipotético como un caso paradigmático de razonamiento default.

El sistema consta de tres conjuntos categorialmente diferentes, que se suponen provistos por el usuario:

F es un conjunto de fórmulas cerradas llamadas 'hechos'. Se supone que F es consistente. Las fórmulas de F se interpretan como verdaderas (no son cuestionables en el sistema).

Δ es un conjunto de fórmulas (posiblemente abiertas) llamadas 'hipótesis posi bles'. Cualquier instancia fundada de un elemento de Δ puede ser utilizada en una

[1]Poole, D, 1988, pag 28.

explicación si resulta consistente con los hechos y restricciones. Los elementos Δ constituyen matrices productoras de instancias.

C es un conjunto (posiblemente vacío) de fórmulas cerradas llamadas 'restric ciones'. Las restricciones constituyen un expediente para bloquear instancias de Δ's en casos particulares no deseados. Los elementos de C no pueden cumplir una función inferencial; solo inhibitoria.

Definicion 1: Un escenario de F,Δ,C es un conjunto $D \cup F$ donde D es un conjunto de instancias fundadas de elementos de Δ tal que $D \cup F \cup C$ es consistente.

Definicion 2: Si x es una fórmula cerrada, una explicación de x a partir de F,Δ, C es un escenario de F,Δ,C que implica (deductivamente) a x.

Es decir, x es explicable a partir de F,Δ,C si existe un conjunto D de instancias fundadas de elementos de Δ tal que:

1) $D \cup F \vdash x$

2) $D \cup F \cup C$ es consistente

Definición 3: Una extensión de F,Δ,C es el conjunto de consecuencias lógicas de un escenario maximal de F,Δ,C.

El siguiente teorema conecta la noción de pertenencia a una extensión con la de ser deducible a partir de un escenario consistente con C:

Teorema 1: x es explicable si y sólo si x está en alguna extensión.[2]

Con el objetivo de investigar propiedades formales de la relación de explicabilidad definida por Poole, veremos en la próxima sección algunas propiedades deseables de la relación de consecuencia no-monótona.

3. Formas debilitadas de monotonía: monotonía cauta y monotonía racional.

La búsqueda de propiedades positivas que permitieran caracterizar a los sistemas no-monótonos, dió origen a la definición de formas más débiles de monotonía para la relación de consecuencia.

La propiedad de monotonía (M), como se sabe, es una propiedad que caracteriza a la relación de consecuencia deductiva[3]. Establece, en su versión finitaria, que si una fórmula es consecuencia de un conjunto Γ de fórmulas, entonces es consecuencia de cualquier superconjunto de Γ. Simbólicamente:

Si $\Gamma \vdash x$, entonces $\Gamma \cup \{y\} \vdash x$

dónde Γ es un conjunto de fórmulas, x e y son fórmulas, y $\vdash$ es la relación de consecuencia deductiva.

D. Gabbay, en su intento por encontrar propiedades que la inferencia no-monótona pudiera cumplir, propuso la monotonía cauta (MC) como propiedad deseable de una relación de consecuencia. Esta propiedad, en su formulación finitaria, establece que no se pierden consecuencias al agregar a un conjunto Γ de fórmulas, una fórmula que es consecuencia suya. En símbolos:

---

[2] Véase la prueba del teorema 2.6 en Poole, D. 1988, pag 30.

[3] En Tarski A, 1930, la monotonía en su versión infinitaria aparece como el Teorema 1 de su sistema.

Si $\Gamma \mid\sim x$ y $\Gamma \mid\sim y$, entonces $\Gamma \cup \{y\} \mid\sim x$
donde Γ es un conjunto de fórmulas, x e y son fórmulas y $\mid\sim$ es una relación de consecuencia no-monótona.

Kraus, Lehmann y Magidor, por otro lado, caracterizaron la propiedad de monotonía racional (MR). La misma establece que no se pierden consecuencias al agregar a un conjunto Γ de fórmulas una fórmula cuya negación no es consecuencia de Γ. En símbolos:

Si $\Gamma \mid\sim x$ y $\neg(\Gamma \mid\sim \neg y)$, entonces $\Gamma \cup \{y\} \mid\sim x$[4]
donde Γ, x e y y $\mid\sim$ funcionan como en el caso anterior.

Podríamos plantearnos si existe alguna relación entre MC y MR.

Teorema 2. En sistemas que no admiten inconsistencia, es decir en los que no existe ninguna fórmula x tal que $\mid\sim x$ y $\mid\sim \neg x$, $MR \rightarrow MC$.

Prueba $\Gamma \mid\sim x$ y $\neg(\Gamma \mid\sim \neg y)$, entonces $\Gamma \cup \{ y \} \mid\sim x$ (Pr)
$\Gamma \mid\sim x$ y $\Gamma \mid\sim y$ (Pr)
$\neg(\Gamma \mid\sim \neg y)$ (simplificación y consistencia)
$\Gamma \cup \{y\} \mid\sim x$ (simplificación, producto y MP)

Este es el caso de la lógica no-monótona de Mc Dermott y Doyle. Dicho sistema no cumple MC, y por ende tampoco MR. Para sistemas como éstos se da obviamente: $M \rightarrow MR \rightarrow MC$

Sucesivas restricciones a las condiciones antecedentes, producen propiedades cada vez más laxas.

4. Propiedades formales de la relación de explicabilidad en Poole.

Nos ocuparemos a continuación de la relación de explicabilidad que Poole caracteriza, preguntándonos si cumple o no con los análogos de las propiedades mencionadas en la sección anterior[5]. De modo que en adelante la relación analizada será la de explicabilidad en el sistema de Poole, y no una relación de consecuencia. Por lo tanto entenderemos

$\Delta, F \mid\sim \bar{x}$

como 'x es explicable a partir de los conjuntos Δ y F'.

---

[4]En realidad la formulación original de los autores aparece en forma de regla. Presentada como propiedad metalingüística, adopta la siguiente forma, equivalente a nuestra formulación:

Si $\Gamma \mid\sim x$ y $\neg(\Gamma \cup \{y\} \mid\sim x)$, entonces $\Gamma \mid\sim \neg y$

[5]Si hiciéramos un paralelo entre el presente sistema y un sistema de lógica no-monótona como el de Reiter 'x es explicable a partir de Γ' en Poole corresponde a 'x es consecuencia no-monótona de Γ' en Reiter.

Veremos que en general el sistema de Poole no cumple MC, para lo cuál exhibiremos un contraejemplo. Sin embargo el sistema sí cumple con MC en el caso especial en que Δ,F, C posea una sola extensión.

Teorema 3.La relación de explicabilidad en teorías con una única extensión cumple MC. En símbolos: Si Δ,F,C tiene una única extensión, entonces,

si $\Delta,F,C \mid\!\sim x$ y $\Delta,F,C \mid\!\sim y$,

entonces $\Delta,F,C, \{y\} \mid\!\sim x$

Prueba. Sea Δ,F,C con una única extensión, a saber, E1

E1= Cn ($D \cup F$), siendo $D \cup F$ escenario maximal de Δ,F,C (por def.)

Por hipótesis,

1. $\Delta,F,C \mid\!\sim x$
2. $\Delta,F,C \mid\!\sim y$

Es decir

3. x, y ε E1 (por Teor. 1 y por existencia de una única extensión)

O sea,

4. x ε Cn ($D \cup F$) (por definición)
5. y ε Cn ($D \cup F$) (por definición)

Y por lo tanto

6. $D \cup F \vdash x$
7. $D \cup F \vdash y$
8. $D \cup F \cup C$ es consistente (ya que $D \cup F$ es escenario y, por definición de escenario, $D \cup F \cup C$ es consistente).

Sea ahora Δ, F',C dónde F' es $F \cup \{y\}$

Entonces

9. $D \cup F \cup \{y\} \vdash x$ (de 6. por monotonía de $\vdash$)

Además

10. $D \cup F \cup \{y\} \cup C$ es consistente

Pués si $D \cup F \cup \{y\} \cup C$ fuera inconsistente, siendo (por 8.) $D \cup F \cup C$ consistente, debería cumplirse

11. $D \cup F \vdash \neg y$

Pero 7. y 11. implicarían $D \cup F \cup C$ es inconsistente, contradiciendo 8.

Luego,

12. $\Delta, F, C \{ y\} \mid\!\sim x$ (9. y 10.)

Que es lo que queríamos demostrar.

En síntesis, no se pierden enunciados explicables al extender una teoría T que posee un único escenario, mediante el agregado de enunciados explicables en T.

En teorías con más de un escenario, la relación de explicabilidad no cumple MC. Exhibimos a continuación una contraejemplo:

Δ = default1 mamífero (x) → ¬ vuela (x)

default2 vampiro(x) → vuela (x)

F = hecho1 vampiro (x) → mamífero (x)

hecho2 vampiro (drácula)

En Δ, F podemos explicar

vuela (drácula) (utilizando el default 2)

y también

$\neg$ vuela (drácula) (utilizando el default1)

Es decir, hay una extensión E1 en la que es explicable que Drácula vuela, y otra extensión E2 en la que es explicable que Drácula no vuela.

Si construimos Δ,F' agregando a Δ,F:

hecho3 vuela (drácula)

En la teoría ampliada Δ,F' deja de ser explicable que Drácula no vuela (pués el default1 no puede ya instanciarse con 'drácula')

Ya que Poole permite $\Delta,F,C \mid\sim x$ y $\Delta,F,C \mid\sim \neg x$, no es suficiente probar que la relación de explicabilidad no cumple MC, para probar que no cumple MR.

En efecto, el sistema no cumple MR en general. Pero sí en el caso particular en que la teoría no tiene restricciones.

Teorema 4. La relación de explicabilidad en teorías sin restricciones cumple MR. En símbolos:

Si $\Delta, F \mid\sim x$ y $\neg(\Delta,F \mid\sim \neg y)$, entonces $\Delta,F, \{y\} \mid\sim x$

Supongamos (condiciones antecedentes)

1. $\Delta, F \mid\sim x$ y

2. $\neg (\Delta,F \mid\sim \neg y)$

Por definición de explicabilidad 1. y 2., implican respectivamente:

3. Existe un Di tal que $Di \cup F \vdash x$ y $Di \cup F$ es consistente.

4. No existe un Dj tal que $Dj \cup F \vdash \neg y$ y $Dj \cup F$ es consistente

De 3. se sigue

5. $Di \cup F \cup \{y\} \vdash x$ (x monotonía $\vdash$)

Además,

6. $Di \cup F \cup \{y\}$ es consistente

Pués si $Di \cup F \cup \{y\}$ fuera inconsistente, siendo $Di \cup F$ consistente, debería ocurrir

7. $Di \cup F \vdash \neg y$, contradiciendo 4.

Luego $Di \cup F \cup \{y\}$ es consistente, y de 5. y 6. se sigue (por def. explicabilidad):

8. $\Delta,F, \{y\} \mid\sim x$

Que es lo que queríamos demostrar.

En síntesis, no se pierden enunciados explicables al extender una teoría T que no posee restricciones mediante el agregado de enunciados cuyas negaciones no son explicables en T.

En la versión con restricciones, la relación de explicabilidad no cumple MR. Para mostrarlo, basta el siguiente contraejemplo:

Sea T1 = Δ, F,C con

Δ = default las - aves - vuelan (x) : ave (x) $\rightarrow$ vuela (x)

F = hecho1 ave (tweety)

C = restricción avestruz (x) $\rightarrow$ $\neg$ las - aves - vuelan (x)
En T1 es explicable
vuela (tweety)
Pero si a T1 le agregamos como hecho
hecho2 avestruz (tweety)
obteniendo T2, que Tweety vuela no es explicable en T2.

Ya que no es explicable en T1 que Tweety no es avestruz, puesto que las restricciones sólo cumplen la función de inhibir usos de los defaults, y no pueden cumplir una función inferencial, en T2 estamos agregando un enunciado cuya negación no es explicable en T1. Sin embargo, esta adición hace caer una explicación anterior.

5. Conclusiones.

Poole presenta una propuesta para formalizar el razonamiento revocable que constituye una reconstrucción fácilmente implementable. Posee el atractivo adicional de no recurrir a cambios de lógica: la lógica utilizada es la clásica. Mientras que en otros sistemas para razonamiento revocable la lógica, a través de la relación de consecuencia, carga con el peso de la no-monotonía, en éste esa carga se traslada a la relación de explicabilidad. La inferencia que el sistema recoge formalmente es siempre monótona, y si subyace una inferencia no-monótona en el proceso de instanciación de los defaults, ésta queda relegada a un plano pragmático e informal.

El estudio de las propiedades formales de la relación de explicabilidad que hemos realizado, refleja un comportamiento adecuado del sistema. En efecto, MC se cumple en el caso de teorías con una única extensión, como es deseable y no se cumple en el caso de teorías con múltiples extensiones. Pero este resultado no constituye una falla del sistema de Poole, del modo en que constituye un resultado indeseable en la lógica no-monótona de Mc Dermott y Foyle, por ejemplo. Efectivamente, si sobre la base de información incompleta es posible generar diferentes extensiones que implican enunciados distintos, y hasta posiblemente contradictorios, resulta razonable esperar que el agregar a dicha base un enunciado perteneciente a una de las extensiones en la categoría de hecho incuestionable, pueda provocar la caída de enunciados anteriormente explicables.

Con respecto a MR, el cumplimiento de la misma en sistemas sin restricciones refleja la propiedad deseable de que sólo la incorporación de un enunciado inconsistente con un conjunto de enunciados puede provocar la desaparición de enunciados explicables en dicho conjunto. La violación de la propiedad en sistemas con restricciones, sólo refleja el hecho de que las restricciones son, justamente, enunciados que recogen excepciones y cuya misión consiste expresamente en inhibir usos de los defaults, pero que no pueden cumplir una función inferencial. El agregado de una restricción, o de un enunciado que la active, puede, aunque no contradiga enunciados anteriores, causar legítimamente la pérdida de enunciados explicables.

El sistema de Poole parece cumplir, en todos los casos, con una propiedad aún más débil que MC, a la que podríamos llamar monotonía ultracauta (MUC) (La lógica default de

Reiter no la cumple, y ésta constituye una deficiencia de su sistema). Su formulación sería: sea T1 una teoría con extensiones E1, E2, ... En . Sea T2 = T1$\cup$\{x\}, con x ε Ei(i ε N, 1= i = n) Entonces, para toda extensión Ej de T2, siendo Ej $\neq$ E1, ... , En, se cumple Ei $\subseteq$ Ej.

El estudio de ésta y de nuevas propiedades que resulten útiles para comprender el funcionamiento de sistemas para razonamientos revocables es tarea para el futuro.

6. Bibliografía.

Kraus,S, Lehmann, D y Magidor M: 'Nonmonotonic Reasoning, Preferential Models and Cumulative Logics' En Artificial Intelligence 44 (1990).

Legris, J:'Diferentes criterios para analizar la inferencia revocable: observaciones sobre el punto de vista lógico'. Actas de las Jornadas de Epistemología de las Cs. Ecs. 1995

Lerner, S: 'Algunas propiedades de la Inferencia no-monótona'. Comunicación V Jornadas de Epistemología e Historia de la Ciencia. Córdoba. 1994.

Lombardi, C: 'Razonamiento Revocable: Aplicaciones en el Area Económico-Administrativa e Implementaciones Computacionales'. Informe de Beca. 1996

Makinson, D: 'General Theory of Cumulative Inference'. En Lecture Notes in Artificial Intelligence. Springer- Verlag - 1988

Mc Dermot,D: 'Nonmonotonic Logic II: Nonmonotonic modal Theories'. En Journal of the Association for Computing Machinery. Vol 29, N* 1, 1982.

Dermott,D y Doyle, J: 'Non-monotonic Logic I'. En Artificial Intelligence', 13, 1980

Poole, D:'A Logical Framework for Default Reasoning'.En Artificial Intelligence 36,1988

Poole, D:'A Methology for Using a Default and Abductive Reasoning System'. En Inter national Journal of Intelligent Systems. Vol 5. Wiley & Sons, 1990.

Reiter,R:'A Logic for Default Reasoning'. En Artificial Intelligence 13, 1980.

Tarski, A: Logic, Semantics, Metamathematics. Hackett Publisching Company, 1930