Razonamiento Abductivo en Teorías no Monotónicas

Claudio Delrieux
Universidad Nacional del Sur, Bahia Blanca - ARGENTINA
claudio@acm.org

Resumen

El razonamiento abductivo es aquel tendiente a encontrar explicaciones paraa observaciones sorprendentes o anómalas en algún sentido. Como tal, es utilizado y posee diversas aplicaciones, particularmente en la inteligencia artificial (sistemas de diagnóstico, resolución de problemas, planificación de tareas, aprendizaje, depuración algorítmica de programas, interpretación del lenguaje natural, y muchos otros). Sin embargo, hasta ahora ha recibido escasa atención la incorporación de patrones abductivos de inferencia en teorías no monotónicas. El objetivo de este trabajo consiste en investigar los modelos de integración del razonamiento no monotónico con la inferencia abductiva. Para ello es necesario generalizar los mecanismos de generación, comparación y aceptación de explicaciones en teorías no monotónicas, y encontrar en cada caso los mecanismos de resolución de conflictos y las estrategias computacionales asociadas.

Palabras Clave: Inteligencia Artificial — Representación del Conocimiento y Razonamiento — Razonamiento no Monotónico — Abducción

1 Introducción

En el razonamiento revisable (no monotónico, argumentativo, default, etc.) se utiliza un esquema de inferencia que utiliza premisas no correctas (unsound) de la forma $a \succ\!\!\!- b$ denominadas *reglas derrotables* (defeasible rules). De esta forma se encadenan inferencias al estilo de las deducciones clásicas (llamadas *teorías* en Reiter, Poole, Delgrande [5, 17, 16, 18], o *argumentos* en Loui, Simari-Loui, Vreeswijk) [14, 20, 22]. La forma de estos argumentos es semejante a una deducción clásica, excepto por el hecho de tener este tipo de premisas. Por dicha razón, podríamos mencionar que este razonamiento incluye un mecanismo de inferencia ampliativo. Hasta ahora no se ha considerado en detalle qué ocurre si se incorporan otro tipo de patrones ampliativos de inferencia, especialmente aquellos que provienen de *reglas* no correctas, como por ejemplo la abducción, la inducción

y la analogía[1]. Por ejemplo, una "teoría abductiva" sería una línea de razonamiento en la cual en algún paso de inferencia se utilizó el patrón ampliativo de la abducción.

Esta es una notable carencia, dado que un aspecto central en el razonamiento humano en todas sus expresiones, consiste en su capacidad para construir explicaciones abductivas. Es posible citar un sinnúmero de ejemplos y circunstancias donde el objetivo final del razonamiento es la producción de explicaciones abductivas para los hechos observados. En las ciencias de la computación, particularmente en la inteligencia artificial, podemos encontrar notables aplicaciones de la abducción, por ejemplo en los sistemas de diagnóstico, en la resolución de problemas, planificación de tareas, depuración algorítmica de programas, interpretación del lenguaje natural, y muchos otros. En la teoría de la ciencia también la abducción representa un papel indispensable en la formalización de la epistemología de las ciencias experimentales. En particular, el paradigma hipotético-deductivo, el cual es hegemónico desde hace más de 50 años, puede pensarse como un esquema abductivo de explicación científica [11, 12, 15][2].

El objetivo de este trabajo consiste en investigar los modelos de integración del razonamiento no monotónico con la inferencia abductiva. Para ello es necesario generalizar los mecanismos de generación, comparación y aceptación de explicaciones en el contexto de las teorías no monotónicas, y encontrar en cada caso los mecanismos de resolución de conflictos y las estrategias computacionales asociadas. En la próxima Sección estableceremos una caracterización lógica de la abducción, la cual, en la Sección 3 será generalizada en el contexto de las teorías no monotónicas, en particular en el razonamiento argumentativo. En la siguiente Sección se establecen las condiciones por las cuales esta inferencia abductiva produce explicaciones estables, en particular cuando las observaciones son o bien consecuencia (monotónica o no) del contexto, o bien cuando son hechos novedosos o sorprendentes. En la Sección 5 se muestra de qué manera modificar el razonamiento abductivo para obtener explicaciones para observaciones anómalas.

2 Inferencia abductiva

La abducción ha sido explícita o implícitamente utilizada en diversos sistemas de razonamiento, fundamentalmente en los llamados "sistemas expertos" que fueron moda hace unos 15 años. Estos sistemas en general incorporan un conjunto de conocimiento representado con reglas (al estilo de PROLOG u OPS5), poseen un conjunto previo de hipótesis abducibles (conjeturas, observables o síntomas), e intentan demostrar el cumplimiento de alguna situación hipotética (diagnóstico o conclusión). Se han utilizado muy exitosamente en dominios tan diversos como en el diagnóstico médico, la prospección minera, el asesoramiento financiero, la configuración de arquitecturas computacionales y muchos otros

[1]Una excepción destacable es el trabajo de Boutilier y Bécher [3], donde se presenta una relación entre la abducción y el modelo AGM de cambio de teorías [1]. El modelo se basa en las lógicas proposicionales bimodales CO, y CT4O [2], en las cuales si las observaciones a explicar son aceptadas (son verdaderas en algún mundo posible), entonces la explicación E es preferida en los mundos más plausibles accesibles desde los mundos donde las observaciones son verdaderas. Esta versión del razonamiento abductivo es la primera en poseer la virtud de poder acomodar hechos inconsistentes con la teoría subyacente (observaciones anómalas).

[2]Es notable el reciente interés que está teniendo la temática de la formalización computacional del razonamiento científico [8, 9].

problemas [4, 19, 21, 23]. Una característica atractiva de esta tecnología es que en general existe un formato homogéneo (o protocolo) para poder representar el conocimiento de cualquier dominio. De esa forma, es posible separar nítidamente la representación del conocimiento de su manipulación por medio del razonador. Por ello es que se consigue una arquitectura de software bastante modular, donde por un lado tenemos el *motor de inferencia* que acepta conocimiento genérico en el protocolo establecido, por otro lado tenemos la interfase con el usuario, también genérica y dependiente del motor de inferencia, y por último está la base de conocimientos, particular para cada aplicación, que permite al motor encadenar los razonamientos, generar las explicaciones, obtener el contenido de los mensajes requeridos por la interfase, y manejar las consultas con el usuario.

Pese al relativo éxito de estas arquitecturas de software, es muy poco lo que se ha formalizado respecto a su funcionamiento. Una posible razón de ello es que en muchos casos el motor de inferencia posee implementados procedimientos heurísticos o *ad hoc* que impiden un tratamiento formal del modelo de razonamiento dentro del márco de la lógica. Una excepción a ésto es el trabajo de Konolige donde intenta formalizar un sistema de razonamiento para propósitos de diagnóstico [13]. Las explicaciones abductivas, bajo la suposición de que el contexto es cerrado, son una forma correcta y consistente de encontrar justificaciones para las observaciones.

Podemos referirnos a la abducción como a una relación ternaria

$$\mathcal{T}, h \rightarrow e, \tag{1}$$

donde la relación de inferencia $\rightarrow$ puede ser la consecuencia lógica clásica $\vdash$, alguna relación de consecuencia no monotónica $\mid\sim$, la derivabilidad en programas lógicos, etc. La relación de inferencia no muestra cómo se genera el proceso, por lo que debe existir también un "mecanismo disparador" que haga necesaria la abducción. Como mecanismo disparador de la abducción, podemos también referirnos a dos casos diferentes: el hecho novedoso o sorprendente (cuando $\mathcal{T} \not\rightarrow e$), y la anomalía (cuando $\mathcal{T} \rightarrow \neg e$)[3]. En el primer caso la abducción puede realizarse aún siendo $\rightarrow$ un operador monotónico, pero claramente la abducción de una anomalía no puede realizarse si $\rightarrow$ es monotónico y manteniendo el contexto de $\mathcal{T}$ fijo al mismo tiempo.

La primer condición que debe cumplir una hipótesis h para contar como posible explicación para e es que (dentro del contexto de la teoría $\mathcal{T}$) h implique consistentemente a e. También puede darse el caso de que explicaciones consistentes sean innecesarias porque e misma es consecuencia de $\mathcal{T}$ o de h por separado (explicación no trivial). Muchas explicaciones que pueden cumplir con estos tres requisitos, igual no cuentan como hipótesis a considerar. Por ejemplo, $\mathcal{T} \Rightarrow e$ es una explicación poco razonable. Al mismo tiempo, quisiéramos evitar explicaciones innecesariamente fortalecidas. Por ejemplo, si h es una explicación para e, y h' es consistente con $\mathcal{T}$, entonces $h \wedge h'$ es también una explicación (indeseada) para e. En particular, definiremos como *explicaciónes más generales* (menos específicas), a aquellas explicaciones consistentes no triviales que sean implicadas a partir de $\mathcal{T}$ por cualquier otra explicación consistente no trivial.

DEFINICIÓN 1 **(Explicaciones generales):** *h es* explicación más general *para e dado el contexto $\mathcal{T}$ sí y solo sí*

[3]Esta clasificación coincide con la *expansión* de $\mathcal{T}$ por e, y la *revisión* de $\mathcal{T}$ por e, respectivamente, en la teoría AGM de cambio de creencias [1, 10].

- $\mathcal{T}, h \models e$,
- $\mathcal{T}, h \not\models \bot$,
- $\mathcal{T} \not\models e$,
- $h \not\models e$,
- *Cualquier otra explicación no trivial h' para e dado $\mathcal{T}$ es tal que $\mathcal{T} \models h' \Rightarrow h$.*

Cuando el contexto teórico $\mathcal{T}$ esté claramente definido, diremos simplemente que h permite encontrar una demostración más general *para e.*

La generalidad de una explicación es un criterio de simpleza o minimalidad.

TEOREMA **1** *$\mathcal{T} \Rightarrow e$ es siempre la explicación más general para e dado $\mathcal{T}$.* □

DEMOSTRACIÓN **1** *Sea h' cualquier otra explicación no trivial para e dado $\mathcal{T}$. Por lo tanto $\mathcal{T}, h' \models e$. Aplicando el teorema de la deducción obtenemos $h' \models \mathcal{T} \Rightarrow e$. Aplicando nuevamente vemos que $\models h' \Rightarrow (\mathcal{T} \Rightarrow e)$.*□

Este resultado parece desalentador, pero no hace más que obligarnos a enfocar la busqueda abductiva de explicaciones (como proceso) a un subconjunto restringido del lenguaje. En nuestro caso, como veremos en la siguiente Sección, realizaremos abducciones sobre activadores de reglas derrotables.

3 Abducción en teorías no monotónicas

En esta Sección consideramos la inclusión de una relación de consecuencia abductiva como la vista en la Sección anterior, pero en el contexto del razonamiento no monotónico. Utilizaremos como modelo de razonamiento no monotónico a las *teorías no monotónicas* (al estilo Poole [16]), o *argumentos* (al estilo Vreeswijk [22]).

DEFINICIÓN **2** *Dado un conjunto de conocimiento firme $\mathcal{K}$, el cual en principio contiene solamente las verdades lógicas y matemáticas, (si existe algún elemento adicional se indica en forma explícita), sean*

- $\mathcal{T}$, *la* teoría, *la cual es un conjunto de condicionales derrotables de la forma $A \succ\!\!- B$, donde en general A y B son conjuntos o conjunciones de literales aplicados a una misma tupla de variables libres,*
- E, la evidencia, *la cual es un conjunto de literales de base, que puede servir tanto para activar argumentos en $\mathcal{T}$ como para requerir explicaciones abductivas, o ambas cosas.*

Entonces el contexto *de nuestro razonador es $\mathcal{T} \cup E$, con el conocimiento subyacente $\mathcal{K}$.*

DEFINICIÓN **3** *Dado un contexto $\mathcal{T} \cup E$, sea* $\mathbf{T}$ *el conjunto de todas las instancias de base de $\mathcal{T}$. En ese caso definimos a la* relación de consecuencia revisable $\mid\!\sim$ *de manera que $A \mid\!\sim B$ sí y solo sí existe una secuencia $a_1, a_2, \cdots, a_n$ tal que $a_n = B$, y cada a_i es una instancia de esquema de axioma, o pertenece a A, o se obtiene por medio de* modus ponens *a partir de miembros anteriores de la secuencia. Los elementos de* $\mathbf{T}$ *se utilizan en la relación $\mid\!\sim$ como implicaciones materiales únicamente para la regla de* modus ponens, *pero no se pueden utilizar contrapositivamente, o debilitar su consecuente, o fortalecer su antecedente.*

Una forma más intuitiva de entender la necesidad del operador $\mid\!\sim$ es la de poder expresar las extensiones no monotónicas de una teoría. De esa manera, $(\mathcal{T} \cup E) \mid\!\sim e$ representa el hecho que e pertenece a alguna extensión del contexto formado por la teoría $\mathcal{T}$ y la evidencia E. En este caso, diremos indistintamente que en el contexto *existe una teoría* para e o bien que *existe un argumento* para e.

DEFINICIÓN **4** *Sea un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$. Entonces un* argumento *o teoría para una conclusión c es un subconjunto $T \subseteq \mathbf{T}$ y un subconjunto $E' \subseteq E$ tal que*

- $\mathcal{K} \cup T \cup E' \mid\!\sim c$,
- $\mathcal{K} \cup T \cup E' \not\mid\!\sim \bot$,
- $\nexists T' \subseteq T.(\mathcal{K} \cup T' \cup E' \mid\!\sim c)$.

En estas circunstancias diremos que E' activa *al argumento T. Cuando no existen elementos en $\mathcal{K}$ además de las verdades lógicas, entonces es posible omitirlo.*

Es posible observar, entonces, que la activación de las reglas o condicionales derrotables se realiza por medio de una instancia de base del antecedente (conjunto de literales), algo que en el razonamiento abductivo no poseemos. Sin embargo, es posible utilizar un mecanismo similar, que a partir de una instancia de base del consecuente de un condicional derrotable encuentre el antecedente que corresponde.

DEFINICIÓN **5** *Sea un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$. Una* observación *$o \in E$ es un literal de base que puede sustituirse por el consecuente de alguna regla en $\mathcal{T}$.*

DEFINICIÓN **6** *(Abducción en teorías no monotónicas).*
Dada una teoría no monotónica $\mathcal{T}$ compuesta por un conjunto de reglas derrotables. Entonces de una premisa (observación) o se infiere la explicación h sí y solo sí

- $h \cup \mathcal{T} \not\mid\!\sim \bot$,
- $\mathcal{T} \not\mid\!\sim o$,
- $h \not\mid\!\sim o$,
- $\mathcal{T} \cup h \mid\!\sim o$,

- *Todo otro conjunto de sentencias h' que satisface los puntos anteriores. es tal que $h' \cup \mathcal{T} \not\sim h$.*

h es más general si es un activador para una regla en $\mathcal{T}$ que tiene a o como consecuente, y es más específica si alguno de los literales de h no es sustituible por el consecuente de ninguna regla en $\mathcal{T}$ (ver Definición 1).

Este mecanismo de abducción en teorías no monotónicas sirve al propósito de encontrar explicaciones para las observaciones existentes en el contexto. Puede darse el caso de que dicha observación esté en efecto justificada por un argumento, es decir, pertenece a una extensión de la teoría cuya base se encuentra en E. En otras situaciones, la observación será "sorprendente" en el sentido de que no tiene una justificación en el contexto dado. En esos casos, el objetivo de nuestro sistema será encontrar una *explicación* para tal observación. La explicación, de existir, consistirá en nuevos literales de base que dispararán argumentos no derrotados cuya conclusión justifica la observación novedosa. Las siguientes definiciones permiten formalizar estos aspectos.

EJEMPLO **1** *Supongamos la teoría.*

$$\begin{array}{ll} \mathcal{T}=\{ & a(X) \succ\!\!- b(X), \\ & b(X) \succ\!\!- c(X), \ \} \end{array}$$

y la observación $o = c(t)$. En ese caso la explicación más general es $b(t)$ y la más específica es $a(t)$.

4 Teorías abductivas estables

De acuerdo a las definiciones en la Sección anterior, dado un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$, y una observación $o \in E$, entonces pueden existir explicaciones H más específicas para o cuando existen argumentos para o en este contexto, tales que $H \subseteq E$. Es además fácil ver que la situación conversa no es válida. Considérese el siguiente ejemplo.

EJEMPLO **2** *Supongamos la teoría.*

$$\begin{array}{ll} \mathcal{T}=\{ & a(X) \succ\!\!- b(X), \\ & b(X) \succ\!\!- c(X), \\ & a(X) \succ\!\!- \neg c(X) \ \} \end{array}$$

Si buscamos una explicación más específica para la observación $c(t)$, encontramos que $a(t)$ podría ser tal explicación, dado que $a(t)$ permite derivar $b(t)$ y ésta, a su vez permite derivar $c(t)$. Sin embargo, $c(t)$ no pertenece a extensiones consistentes del contexto $\mathcal{T} \cup \{c(t)\}$, y por lo tanto la suposición de $a(t)$ no genera la explicación deseada.

Por consiguiente, la no localidad en el razonamiento no monotónico nos obliga a encontrar las explicaciones más específicas utilizando abducción más general iterada.

DEFINICIÓN **7** *Sea un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$. Entonces*

1. *Se generan todas las posibles explicaciones abductivas B a partir de $E \cup \mathcal{T}$ (cada elemento $b \in B$ es una posible explicación más general),*

2. *Se generan todas las conclusiones A que pertenecen a extensiones consistentes de $E \cup \mathcal{T}$,*

3. *En caso de contradicción entre algún literal $a \in A$ con algún literal $b \in B$ consideramos que el razonamiento que genera la extensión a la que a pertenece tiene prioridad sobre el literal abducido b.*

4. *Los conjuntos de conclusiones firmes C son aquellos miembros de A y de B que quedan firmes en el paso anterior.*

Si se requiere realizar una explicación abductiva iterada, entonces se incorpora un elemento de B (elegido arbitrariamente) al conjunto E, y se aplica nuevamente el proceso.

EJEMPLO **3** *Supongamos contar con la teoría no monotónica siguiente.*

$$
\begin{array}{lll}
\mathcal{T}=\{ & q(X) \succ\!\!\!- \; p(X), & \text{(los Cuáqueros son pacifistas)}, \\
& q(X) \succ\!\!\!- \; rel(X), & \text{(los Cuáqueros son religiosos)}, \\
& r(X) \succ\!\!\!- \; b(X), & \text{(los Republicanos son belicistas)}, \\
& p(X) \succ\!\!\!- \; \neg b(X), & \text{(los pacifistas no son belicistas)}, \\
& b(X) \succ\!\!\!- \; \neg p(X), & \text{(los belicistas no son pacifistas)}, \\
& b(X) \succ\!\!\!- \; dm(X), & \text{(los belicistas apoyan la defensa misilística)}, \\
& b(X) \succ\!\!\!- \; pm(X), & \text{(los belicistas son políticamente motivados)}, \\
& p(X) \succ\!\!\!- \; pm(X) \; \}. & \text{(los pacifistas son políticamente motivados)}.
\end{array}
$$

Si nuestra evidencia E es que $rick$ es políticamente motivado ($pm(rick)$), entonces las dos explicaciones más generales que podemos generar son o bien que $rick$ es pacifista ($p(rick)$) o bien que es belicista ($b(rick)$), ambas incompatibles entre sí.

Iterando el proceso, podemos encontrar explicación para $p(rick)$ en $q(rick)$, y explicación para $b(rick)$ en $r(rick)$. En estas circunstancias, cualquiera de estos dos casos son explicaciones más específicas de $pm(rick)$, y además, incompatibles entre sí.

Una caracterización importante en este proceso consiste en determinar cuándo es posible encontrar una explicación consistente con el contexto.

DEFINICIÓN **8** *Sea un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$. Dada una observación o compatible con el contexto, pero sin justificación en el mismo, diremos que o tiene una* justificación estable *H si existe por lo menos una explicación más específica H para o, que sea consistente con el contexto.*

TEOREMA **2** *El procedimiento de la Def. 7 aplicado exhaustivamente encuentra siempre una justificación estable, si es que ésta existe.*

DEMOSTRACIÓN **2** *Que H sea una justificación estable implica que es una explicación no trivial, más específica, y compatible con el contexto, es decir:*

1. $H, \mathcal{K} \nvdash o$,

2. $H, \mathcal{K}, \mathcal{T}, E \mid\sim o$, *(explicación no trivial),*

3. *Los literales en* H *no ocurren el el consecuente de ninguna regla en* $\mathcal{T}$*, (explicación más específica),*

4. *No es posible a partir de* $H, E, \mathcal{T}$ *generar argumentos para un conjunto* C *de literales tal que* $C, \mathcal{K} \vdash \bot$*, (explicación compatible con el contexto).*

Si existe H *en tales condiciones, entonces es posible construir un árbol que tiene a* o *como raíz, a los literales de* E *y* H *como hojas, cada arco es una instancia de base de una regla en* $\mathcal{T}$*, y cada nodo interior vincula un conjunto de consecuentes* S *de instancias de base de una regla en* $\mathcal{T}$*, con el activador* r *de otra regla en* $\mathcal{T}$*, de modo que* $\mathcal{K}, S \vdash r$*.*

En estas condiciones queremos probar que si existe un conjunto H *con estas propiedades, entonces el procedimiento lo encuentra, y si el procedimiento fracasa, es porque no existe tal* H*. El procedimiento genera una recorrida por niveles del árbol (*breadth-first*) en el paso 1, filtrando los descendientes incompatibles con el contexto en el paso 2. Por lo tanto, es posible realizar la demostración por inducción sobre la recorrida de dicho árbol.*

Cuando el procedimiento se aplica por primera vez (sobre la observación o*), en el paso 1 se generan los descendientes* B *de* o*, y en el paso 2 se generan todas las conclusiones* A *justificadas a partir del contexto. Los literales en* A *tienen precedencia sobre* B*. Si todos los literales en* B *fueron filtrados, entonces no hay justificación estable y el procedimiento fracasa. Si existe por lo menos un nodo descendiente que no fue filtrado, entonces se hace el mismo análisis sobre cada subárbol generado por cada uno de dichos nodos (el siguiente nivel del árbol).*

Entonces por inducción es posible ver que o bien todos los respectivos descendientes fueron eventualmente filtrados (no hay justificación estable), o bien existe un conjunto de nodos H *que es descendiente de* o*, y ninguno de dichos nodos es instancia del consecuente de ninguna regla en* $\mathcal{T}$*. En esta situación,* H *forma parte junto con* E *de las hojas del árbol que tiene a* o *como raíz, y es por lo tanto la justificación estable buscada.* □

EJEMPLO 4 *Supongamos la siguiente teoría.*

$$
\begin{array}{lll}
\mathcal{T}=\{ & ll(T) \succ\!\!\!- cm(T), & \text{(si llueve, la calle está mojada),} \\
& ll(T) \succ\!\!\!- pm(T), & \text{(si llueve, el patio está mojado),} \\
& ll(T) \succ\!\!\!- \neg s(T), & \text{(si llueve, no hay sol),} \\
& s(T) \succ\!\!\!- \neg ll(T), & \text{(si hay sol, no llueve),} \\
& a(T) \succ\!\!\!- pm(T), & \text{(si prendieron los aspersores, el patio está mojado),} \\
& s(T) \wedge c(T) \succ\!\!\!- a(T), & \text{(si hubo sol y calor, funcionaron los aspersores),} \\
& pm(T) \succ\!\!\!- zm(T), & \text{(si el patio está mojado, se mojan los zapatos),} \\
& cm(T) \succ\!\!\!- zm(T) \ \}. & \text{(si la calle está mojada, se mojan los zapatos).}
\end{array}
$$

En esta situación, supongamos que observamos que tenemos los zapatos mojados $(zm(hoy))$*, y creemos recordar que hubo sol* $(s(hoy))$*. ¿Qué puede concluirse al respecto?*

Las explicaciones abductivas más generales son $cm(hoy)$ *y* $pm(hoy)$*. Por el momento cualquiera de las dos es compatible con las observaciones y los argumentos generados a partir de ellas. Por lo tanto ambas pasarían a pertenecer al conjunto* C*. Al iterar el proceso de abducción, para encontrar una explicación posible para alguno de los miembros de* C *nos encontramos con la siguiente situación. La búsqueda de una explicación para*

$cm(hoy)$ queda bloqueada, porque la única explicación es $ll(hoy)$, que es incompatible con un argumento generado a partir del contexto y la teoría (basado en que $s(hoy)$).

La búsqueda de una explicación para $pm(hoy)$ tiene dos posibilidades. La primera también es $ll(hoy)$ y también queda bloqueada, mientras que la segunda, $a(hoy)$ es compatible con el contexto. Siguiendo la iteración, entonces, asimilamos a $s(hoy) \wedge c(hoy)$ como única justificación estable que se puede encontrar para $zm(hoy)$.

Es decir, hay sol y tenemos los zapatos mojados. Conjeturamos que funcionaron los aspersores, el patio se mojó y ello mojó nuestros zapatos. Aplicando una nueva abducción, podríamos conjeturar que la causa del funcionamiento de los aspersores fue que, además de que hubo sol —algo confirmado en la experiencia— también hubo calor.

5 Abducción de observaciones anómalas

Un corolario de los resultados de la Sección anterior es que explicar una observación incompatible con el contexto implica modificar el contexto. Es decir, puede darse el caso en el que una observación no pueda explicarse sin que *algún* elemento en el contexto deba ser modificado. En este caso, la observación es "anómala" en algún sentido, dado que directa o indirectamente produce una situación contradictoria en el estado de creencias del sistema, dado el contexto.

Tal vez el ejemplo más sencillo de observaciones anómalas, el cual, sin embargo, contiene todos los elementos necesarios para nuestro análisis, es el siguiente.

EJEMPLO **5** *Nuestra teoría $\mathcal{T}$ tiene una sola regla $a(X) \succ\!\!- b(X)$, la evidencia E contiene al literal $a(t)$, y a la observación o a justificar es $\neg b(t)$.*

En esta situación, las únicas elecciones posibles consisten en rechazar la observación por ser anómala, o bien rechazar la evidencia por incompatible, o bien asumir que t es una excepción a la regla en $\mathcal{T}$. Podríamos afirmar que la Def. 7 está en sintonía con la primera de estas posibilidades, lo cual en ciertos casos puede ser razonable.

Sin embargo, es importante discutir qué debería hacerse si la observación anómala es lo suficientemente persistente como para requerir una sensata revisión del contexto. Este tipo de situaciones son en realidad moneda corriente en el razonamiento científico, donde el cambio en las teorías científicas ocurre cuando la evidencia experimental no es compatible con el conocimiento establecido. Es importante considerar el siguiente resultado.

TEOREMA **3** *Sea un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$, y una observación anómala $o \in E$. En estas condiciones, o bien o no puede explicarse, o bien hay que modificar el contexto (abandonar algún literal en E o bloquear alguna instancia de regla activada en $\mathcal{T}$).*

DEMOSTRACIÓN **3** *Que o sea anómala implica que no existe una justificación estable. Si no estamos dispuestos a modificar el contexto, entonces estamos en las condiciones del Teorema 2 y por lo tanto no es posible encontrar una explicación para o.*

Si hemos encontrado una explicación para o, quiere decir que hemos encontrado

un conjunto de literales H tales que junto con instancias de reglas en $\mathcal{T}$ nos permiten generar un argumento no derrotado para o. Como o es anómala, quiere decir entonces que dentro del contexto era posible construir por lo menos un argumento derrotador para cada argumento para o. Por lo tanto, por lo menos uno de esos argumentos derrotadores ha sido descartado para dar paso al argumento para o. La única manera de descartar un argumento consiste o bien en eliminar algún literal de su base en E, o bien en bloquear alguna instancia de las reglas en $\mathcal{T}$ que se utilizan para generarlo. □

Para poder encontrar un mecanismo que permita explicar observaciones anómalas, entonces, debemos modificar nuestro procedimiento, probablemente modificando la prioridad que se otorga a los literales abducidos respecto de los literales generados por argumentos.

DEFINICIÓN **9** *(Abducción de observaciones anómalas)*

Sea un contexto $\mathcal{T} \cup E$ con un conocimiento subyacente $\mathcal{K}$. Entonces

1. *Se generan todas las posibles explicaciones abductivas B a partir de $E \cup \mathcal{T}$ (cada elemento $b \in B$ es una posible explicación más general),*
2. *Se generan todas las conclusiones A justificadas argumentativamente a partir de $E \cup \mathcal{T}$,*
3. *En caso de desacuerdo entre algún literal $a \in A$ con algún literal $b \in B$,*
   (a) *Las inferencia abductiva para b derrota a todo argumento T para a, o bien,*
   (b) *Sea $\mathcal{T}_T \subseteq \mathcal{T}$ el subconjunto de reglas utilizadas para la generación de cada argumento T para a. Entonces existe por lo menos una regla en $\mathcal{T}_T$ que es sintácticamente bloqueada para la instancia de base con la que fue activada en T*[4].
4. *Los conjuntos de conclusiones firmes C son aquellos miembros de A y de B que no fueron derrotados.*

El caso 3a corresponde a diseñar un sistema en el cual la inferencia abductiva de las observaciones derrota a la evidencia, mientras que el caso 3b corresponde a diseñar un sistema en el cual la contradicción entre evidencia y observaciones derrota a las reglas.

EJEMPLO **6** *(ejemplo 5 revisitado) Nuestra teoría $\mathcal{T}$ tiene una sola regla $a(X) \succ\!\!- b(X)$, la evidencia E contiene $a(t)$ y $\neg b(t)$. En estas condiciones, podemos utilizar el caso 3a de la definición más arriba, y concluír que $a(t)$ debe ser erróneo. En cambio, si utilizamos el caso 3b, entonces aceptamos la veracidad de $a(t)$ pero postulamos que la regla $a(X) \succ\!\!- b(X)$ tiene una excepción para el individuo t.*

[4]Una regla derrotable puede bloquearse sintácticamente para un caso particular si explícitamente impedimos que su activador genere el consecuente correspondiente. Esto significa que la regla puede seguir siendo utilizada para otros activadores.

EJEMPLO **7** *(Ej. 3 revisitado).*
Supongamos contar con la teoría no monotónica $\mathcal{T}$ del Ejemplo 3. En dicha situación, si nuestra evidencia E es que rick es Cuáquero ($q(rick)$) y que apoya la defensa misilística ($dm(rick)$). La explicación más general para $dm(rick)$ es que rick es belicista, lo cual es contradictorio con un argumento generado a partir de la evidencia de que rick es Cuáquero.

Si en este punto detenemos la abducción, y utilizamos el caso 3a para resolver el conflicto, entonces rechazamos la veracidad de que rick sea Cuáquero, por lo que luego podemos encontrar explicaciones más específicas para su belicismo (en particular, que rick es Republicano).

En cambio, si uilizamos el caso 3b en este punto, entonces asumimos que rick es una excepción a alguna de las reglas involucradas (o bien es un Cuáquero no pacifista, o bien es un "pacifista belicista").

6 Conclusiones

En este trabajo nos hemos ocupado del problema de incorporar la inferencia abductiva en el razonamiento no monotónico. Nuestra propuesta consideró un tratamiento igualitario entre la inferencia no monotónica con reglas derrotables y la inferencia ampliativa. En el primer caso, la conclusión es tentativa porque se utiliza como premisa una regla derrotable, la cual puede tener excepciones. En el segundo caso, la conclusión también es tentativa, pero porque el esquema de inferencia mismo puede tener excepciones. Se consideraron en particular dos mecanismos de integración entre el razonamiento no monotónico y la abducción, dependiendo de que la observación a justificar sea anómala o meramente sorprendente.

El desarrollo de estos sistemas de razonamiento realizó cierta cantidad de elecciones. Las mismas expresan una particular relación de compromiso entre las distintas ventajas y problemas que puede acarrear adoptar uno u otro punto de vista. Por dicha razón concluimos que el desarrollo de un sistema de razonamiento con objetivos tan amplios debe seguir un proceso de diseño basado en las características y especificaciones que se desea que posea o que no posea.

Referencias

[1] C.Alchourón, P. Gärdenfors, y D. Makinson. On The Logic of Theory Change. *J. Symbolic Logic*, 50(2):510–530, 1985.

[2] Craig Boutilier. Conditional Theories of Normality: a Modal Approach. *Artificial Intelligence*, 68(1):87–154, 1994.

[3] C. Boutilier y V. Becher. Abduction as Belief Revision. *Artificial Intelligence*, 77(1):143–166, 1996.

[4] B. Buchanan y E. Shortlife. *Rule-Based Expert Systems.* Addison-Wesley, 1984.

[5] James Delgrande. An Approach to Default Reasoning Based on a First-Order Conditional Logic: Revised Report. *Artificial Intelligence*, 33:105–130, 1987.

[6] Claudio Delrieux. Computational Theory of Science: Implementing Research Programmes. In *Proceedings of the IC-AI 2000 Conference*, pages 775–783, CSREA Press, ISBN 1-892512-2, 2000.

[7] Claudio Delrieux. The Rôle of Defeasible Reasoning in the Modelling of Scientific Research Programmes. In *Proceedings of the IC-AI 2001 Conference*, pages 861–868, CSREA Press, ISBN 1-892512-81-5, 2001.

[8] Claudio Delrieux (editor). Proceedings of the First Workshop on Computer Modelling of Scientific Reasoning and Application. CSREA Press, ISBN 1-892512-73-4. `www.lip.uns.edu.ar/cmsra`, 2001.

[9] Claudio Delrieux (editor). Proceedings of the Second Workshop on Computer Modelling of Scientific Reasoning and Application. CSREA Press, ISBN 1-892628-23-7. `www.lip.uns.edu.ar/cmsra`, 2002.

[10] Peter Gärdenfors. *Knowledge in Flux*, MIT Press, Cambridge, 1988.

[11] C. Hempel y P. Oppenheim. The Logic of Explanation. *Phil. of Science*, 15:135–175, 1948.

[12] Gregorio Klimovsky. *Las Desventuras del Conocimiento Científico.* A-Z Editora, Buenos Aires, Argentina, 1995.

[13] Kurt Konolige. Abduction versus Closure in Causal Theories. *Artificial Intelligence*, 53(2-3):255–272, 1992.

[14] Ronald P. Loui. Defeat Among Arguments: A System of Defeasible Inference. *Computational Intelligence*, 3(3), 1987.

[15] Ernest Nagel. *The Structure of Science.* Basic Books, New York, 1961.

[16] David Poole. A Logical Framework for Default Reasoning. *Artificial Intelligence*, 36(1):27–47, 1988.

[17] David L. Poole. On the Comparison of Theories: Preferring the Most Specific Explanation. En *Ninth International Joint Conference on Artificial Intelligence*, pp. 144–147, Morgan Kaufmann Los Altos, CA, 1985.

[18] Raymond Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13(1,2):81–132, 1980.

[19] E. Shortlife. *Computer-based Medical Consultations: MYCIN.* Elsevier, New York, 1976.

[20] G. Simari y R. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53(2-3):125–158, 1992.

[21] M. Stefick, J. Aikiens, R. Balzer, J. Benoit, L. Birnbaum, F. Hayes-Roth, y E. Sacerdoti. The Organization of Expert Systems. *Artificial Intelligence*, 18(2):135–173, 1982.

[22] G. A. W. Vreeswijk. Abstract Argumentation Systems. *Artificial Intelligence*, 90(2):225–279, 1997.

[23] D. Watermann. *A Guide to Expert Systems.* Addison-Wesley, Reading, MA, 1986.