Modelos Formales del Razonamiento Científico

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

1 Introducción

El objetivo de esta línea de trabajo consiste en formalizar el razonamiento científico utilizando un lenguaje lógico. Dentro del método científico, especialmente dentro de las ciencias experimentales, se estableció un conjunto de procedimientos que permiten llevar adelante la explicación y predicción de fenómenos y la generación, corroboración y refutación de teorías. La formalización de estos procedimientos constituye uno de los principales objetos de estudio de la teoría de la ciencia. El conocimiento científico se ordena y configura en estructuras complejas. Las unidades de organización más destacadas dentro de esta estructura son las *teorías.* Las teorías científicas tienen la función de establecer conexiones sistemáticas dentro de un aspecto de la realidad. De ese modo es posible la inferencia de determinados hechos a partir de otros. Si el *tipo* de conocimiento que constituye una teoría científica fuese conocimiento verdadero justificado y el método científico fuese deductivamente válido, entonces no habría diferencia entre las teorías científicas y las lógicas. Sin embargo, las teorías científicas involucran tipos de conocimiento cuya justificación es problemática, y mecanismos de inferencia ampliativos. Por lo tanto estas teorías no tienen el *status* de ser deductivamente válidas.

Hempel [3] propone que la lógica de la predicción y de la explicación proceden según un mismo esquema $\mathcal{T} \vdash e$, donde $\mathcal{T}$, el *explanans* es un conjunto de leyes, y e, el *explanandum* es el fenómeno o hecho a explicar. La sistematización por medio de este esquema se denominó *paradigma hipotético-deductivo* o deductivo-nomológico, dado que el *explanans* constituye una pieza de conocimiento hipotético, del cual se debe deducir la evidencia. El procedimiento de inferir el *explanans* no puede ser deductivo, es decir, $\mathcal{T}$ nunca puede ser *verdadera.* Una conclusión, señalada por Popper, es que las teorías científicas no se *verifican* sino que se *refutan.* Dicho de otra forma, no existe evidencia posible que garantice la verdad lógica de una teoría, pero una sola predicción o explicación incorrecta -aunque sea frente a una cantidad enorme de casos correctos- sirve para mostrar que una teoría es falsa.

Una propuesta pragmáticamente más adecuada que el paradigma H-D fue realizada por Lakatos en sus *programas de investigación* [4]. Lakatos observa que una teoría no puede ser absolutamente verdadera. Es más, una teoría puede ser falsa pero tener consecuencias verdaderas y operacionales. Por esta razón los programas de investigación reales permanecen abiertos y sujetos al cambio y la evolución. Esto determina que un programa pueda sobrevivir sin que por ello sea verdadero. Los programas de investigación son estructuras que incluyen a las teorías científicas, integrándolas con un conjunto de procedimientos de inferencia y que poseen un conjunto periférico de hipótesis auxiliares. El *núcleo* de un programa es un conjunto de conocimiento que se considera central, y que define la teoría como tal. Este núcleo es el conjunto de conocimientos (leyes, generalizaciones o postulados) que determina la identidad de la teoría y por consiguiente del programa mismo. El núcleo, por lo tanto, se considera definitivo, y el resto de la estructura del programa opera de modo tal de protegerlo de la refutación. Esta protección consiste básicamente en implementar un *cinturón protector* de hipótesis auxiliares, que impiden que el núcleo sea refutado. De esa forma, existen dos procedimientos heurísticos para confrontar a la teoría con la evidencia de un resultado experimental e. Mientras no ocurran refutaciones, entonces se aplica la heurística positiva, en la cual se busca una sistematización del cinturón protector (como consecuencia del núcleo duro o por medio de nuevas leyes). Si el resultado e no es correctamente predicho o explicado por la teoría, entonces se aplica la heurística negativa, que consiste en encontrar una hipótesis c particular al caso e tal que de la teoría aumentada con c se siga e. Si dicha hipótesis no es compatible con el resto de la teoría, entonces algo en la misma deberá corregirse.

2 Formalizando los *programas* de Lakatos

En esta Sección propondremos una formalización de los *Programas*, en términos de los sistemas de razonamiento no monotónicos. También esbozaremos algunas ideas tendientes a comparar entre sí diferentes *programas* en función de algunas características ventajosas, como ser el progreso empírico y el éxito relativo, dejando planteadas otras posibilidades como por ejemplo el poder sistematizador, o la menor cantidad de hipótesis *ad hoc*. Esta formalización presupone que el conocimiento válido $\mathcal{K}$ es conocido y universalmente aceptado por todos los programas, al igual que la evidencia E. Utilizaremos como mecanismo de inferencia subyacente a un sistema abstracto de razonamiento argumentativo [5, 6]. Es decir, las predicciones o explicaciones en los programas se obtienen construyendo argumentos.

Definición 1 *Dado un conjunto $\mathcal{K}$ de conocimiento válido y un conjunto E de evidencia (literales de base), un* programa de investigación $P = \langle \mathcal{T}, C \rangle$ *consta de una teoría no monotónica $\mathcal{T}$ y un cinturón protector C. La teoría $\mathcal{T}$ está representada por medio de un conjunto de condicionales derrotables $a(X) \rightarrowtail b(X)$, y C por medio de literales de base que representan las hipótesis* ad hoc. *Un* programa de investigación $P = \langle \mathcal{T}, C \rangle$ predice *(o explica) a un literal de base e cuando existe un argumento para e a partir de $\mathcal{K}, \mathcal{T}, C$ y E.*

Es importante recordar que la existencia de un argumento para un literal no es condición suficiente para asumir que la teoría argumentativa acepta a dicho literal como conclusión, dado que pueden existir argumentos en contra que lo derroten. Esto es análogo a lo que ocurre en un programa de investigación, donde el programa puede realizar predicciones para un observable, pero también predicciónes contradictorias con el mismo observable. Por lo tanto deberíamos repasar cuáles son las "actitudes proposicionales" que la teoría argumentativa nos permite asumir respecto del literal e.

Definición 2 *Dado un programa P y un literal de base e, se pueden dar los casos:*

	Definitiva	*Revisable*
Aceptar e	*Argumentos solamente para e*	*Argumentos tanto para e como para $\neg e$ pero existe un argumento no derrotado para e*
Aceptar $\neg e$	*Argumentos solamente para $\neg e$*	*Argumentos tanto para e como para $\neg e$ pero existe un argumento no derrotado para $\neg e$*
Indeciso	*No hay argumentos ni para e ni para $\neg e$*	*Argumentos tanto para e como para $\neg e$ pero no existen argumentos no derrotados*

Dado un programa P y un literal de base e el cual se confirma en la experiencia, se puede dar solamente uno de entre los siguientes casos:

Predicción de P	*Status de P luego de la confirmación de e*
Aceptación definitiva de e	*P se confirma*
Aceptación revisable de e	*P se confirma parcialmente*
Aceptación definitiva de ¬e	*Anomalía grave*
Aceptación revisable de ¬e	*Anomalía parcial*
Indecisión definitiva	*Hecho sorprendente*
Indecisión revisable	*"Laguna"*

EJEMPLO 1 *Supongamos que nuestro conocimiento, evidencia, y suposiciones acerca de las propiedades de algunos animales es el siguiente:*

$\mathcal{K} = \{ping\ddot{u}ino(X) \Rightarrow ave(X)\}$.
$\mathcal{T}_1 = \{ave(X) \succ\!\!- vuela(X), ping\ddot{u}ino(X) \succ\!\!- \neg vuela(X), en\text{-}avi\acute{o}n(X) \succ\!\!- vuela(X)\}$.
$E = \{ping\ddot{u}ino(Opus), ave(Pint\acute{\imath}n), ave(Tweety), ping\ddot{u}ino(Schmuck), en\text{-}avi\acute{o}n(Schmuck)\}$.

Por el momento, nuestro programa no requiere hipótesis auxiliares $C_1 = \{\}$. En estas condiciones, las predicciones que hace nuestro programa $P_1 = \langle \mathcal{T}_1, C_1 \rangle$ con el conocimiento subyacente $\mathcal{K}$ y la evidencia E se basan en lo siguiente:

1. *Existe un argumento para $vuela(Opus)$ (por ser ave), pero hay otro argumento para $\neg vuela(Opus)$ (por ser pingüino) que es más específico y que lo derrota. Por lo tanto P_1 predice $\neg vuela(Opus)$.*

2. *Existe un argumento para $vuela(Tweety)$, por lo tanto P_1 predice dicho literal.*

3. *Existe un argumento para $vuela(Pint\acute{\imath}n)$, por lo tanto P_1 predice dicho literal.*

4. *Existe un argumento para $vuela(Schmuck)$ (por estar en avión) y otro argumento para $\neg vuela(Schmuck)$ (por ser pingüino). Ambos argumentos se derrotan mutuamente, por lo que P_1 no hace predicciones al respecto.*

Supongamos ahora que realizamos observaciones acerca de los individuos voladores, encontrando la siguiente evidencia:

$E_r = \{\neg vuela(Opus), \neg vuela(Tweety), vuela(Gringa), vuela(Pint\acute{\imath}n), vuela(Schmuck)\}$.

En dicho caso, podemos ver que nuestro programa P_1 tiene el siguiente status frente a las nuevas observaciones: con $\neg vuela(Opus)$ tiene una confirmación parcial, con $\neg vuela(Tweety)$ tiene una anomalía grave, con $vuela(Gringa)$ tiene una observación sorprendente, con $vuela(Pint\acute{\imath}n)$ tiene una confirmación, y con $vuela(Schmuck)$ tiene una laguna.

Por el momento, no estableceremos diferencias entre confirmaciones y confirmaciones parciales, lagunas y hechos sorprendentes, y anomalías parciales o graves. La distinción será importante más adelante, cuando tratemos en la próxima Sección de formalizar algunos aspectos de la dinámica de los programas, es decir, cuándo y cómo el programa debe modificarse para adaptarse a las nuevas observaciones. De esa forma, denominaremos V al conjunto de casos donde la evidencia confirma total o parcialmente al programa P, I las indeterminaciones o casos sorprendentes o lagunas para P, y F los casos refutatorios o anomalías parciales o graves. En el ejemplo anterior, podemos ver que $V_1 = \{\neg vuela(Opus), vuela(Pint\acute{\imath}n)\}$, $I_1 = \{vuela(Gringa), vuela(Schmuck)\}$, $F_1 = \{\neg vuela(Tweety)\}$.

DEFINICIÓN **3** *Un programa P_1* es empíricamente más progresivo *que otro P_2 siempre que tenga estrictamente mayor cantidad de casos confirmatorios. Es decir,*

$$P_1 \prec_{prog} P_2 \Leftrightarrow V_2 \subset V_1.$$

P_1 es más exitoso que P_2 si cada caso confirmatorio de P_2 es también confirmatorio de P_1, cada caso refutatorio de P_1 es también refutatorio de P_2, pero hay por lo menos un caso confirmatorio de P_1 que no es confirmatorio de P_2 o por lo menos un caso refutatorio de P_2 que no es refutatorio de P_1. Es decir,

$$P_1 \prec_{ex} P_2 \Leftrightarrow V_2 \subseteq V_1, F_1 \subseteq F_2, \exists v \in V_1 - V_2 \vee \exists f \in F_2 - F_1$$

EJEMPLO **2** *Supongamos estar en una situación similar a la del Ejemplo 1. A partir de la evidencia previa E_d del ejemplo, un programa P_2 edificado a partir una segunda teoría $\mathcal{T}_2$ realizará las siguientes predicciones:*

$\mathcal{T}_2 = \{ave(X) \succ\!\!- vuela(X), ping\ddot{u}ino(X) \succ\!\!- \neg vuela(X)\}.$

1. *Existe un argumento para $vuela(Opus)$ (por ser ave), pero hay otro argumento para $\neg vuela(Opus)$ (por ser pingüino) que es más específico y que lo derrota. Por lo tanto P_2 predice $\neg vuela(Opus)$.*
2. *Existe un argumento para $vuela(Tweety)$, por lo tanto P_2 predice dicho literal.*
3. *Existe un argumento para $vuela(Pintín)$, por lo tanto P_2 predice dicho literal.*
4. *Existe un argumento para $\neg vuela(Schmuck)$ (por ser pingüino), por lo que P_2 predice dicho literal.*

Si luego nos confrontamos a las mismas observaciones

$E_r = \{\neg vuela(Opus), \neg vuela(Tweety), vuela(Gringa), vuela(Pintín), vuela(Schmuck)\},$

podemos ver que el nuevo programa P_2 tiene con $\neg vuela(Opus)$ una confirmación parcial, con $\neg vuela(Tweety)$ tiene una anomalía grave, con $vuela(Gringa)$ tiene una observación sorprendente, con $vuela(Pintín)$ tiene una confirmación, y con $vuela(Schmuck)$ tiene otra anomalía grave. Entonces $V_2 = \{\neg vuela(Opus), vuela(Pintín)\}$, $I_2 = \{vuela(Gringa)\}$, $F_2 = \{\neg vuela(Tweety), vuela(Schmuck)\}$.

En estas condiciones, si bien P_1 y P_2 son igualmente progresivos, dado que coinciden sus casos confirmatorios, P_2 es menos exitoso que P_1 porque existe por lo menos una observación ($vuela(Schmuck)$) que es anómala para P_2 pero no para P_1 .

3 La dinámica de los programas

En la Sección anterior identificamos seis posibles resultados de un programa P confrontado a un resultado experimental e_r. En esta Sección trataremos de dilucidar cuál debería ser la estrategia de un programa en función del status que asume frente a nuevos resultados experimentales. Si el programa se confronta con un hecho sorprendente, entonces debería recurrir a algún mecanismo de inferencia ampliativa para poder explicarlo (en este punto es donde se requiere de un contexto de descubrimiento). En cambio, cuando los resultados son anómalos, entonces es necesario aplicar la heurística negativa, es decir, tratar de modificar las hipótesis periféricas del mismo para defender el núcleo de la refutación o para mantener progresivo al programa. Cuando existe confirmación total o parcial debería aplicarse la heurística positiva, es decir, tratar de aplicar el núcleo del programa para sistematizar también el cinturón protector, y de esa manera reducir la cantidad de hipótesis *ad hoc*. Finalmente, cuando el programa está indeterminado por tener una laguna, es decir, si el programa generó teorías o argumentos no definitivos tanto a favor como en contra de la nueve evidencia, entonces es discutible si alguna de las dos heurísticas debería ser aplicada. Una manera de asimilar la importancia de los mecanismos de inferencia que presentamos en los anteriores Capítulos consiste en verlos como posibles formalizaciones de estas estrategias heurísticas de un programa cuando se confronta con una alguno de los casos mencionados.

EJEMPLO **3** *Supongamos estar en la misma situación que en el Ejemplo 1 y deseamos desde el programa* P_1 *encontrar explicación al hecho sorprendente* $vuela(Gringa)$. *Una estrategia para explicar un hecho sorprendente* e *consiste en encontrar por abducción [1, 2] algún hecho particular* h *el cual, junto con la teoría, permita inferir* e. *En este caso tenemos dos posibles explicaciones más específicas*[1] *encontradas por abducción:* $ave(Gringa)$ *y* $en\text{-}avión(Gringa)$, *hipótesis que luego habrá que corroborar por la experiencia. Mientras que estas hipótesis no estén corroboradas, cualquiera de las dos (o ambas) pueden pasar a formar parte del cinturón protector* C.

EJEMPLO **4** *Siguiendo nuevamente con el Ejemplo 1 desde el programa* P_1. *Tenemos en* $\neg vuela(Tweety)$ *una anomalía grave. Sin embargo, podemos encontrar una explicación abductiva que corrija la anomalía. En efecto, si agregamos la hipótesis* $pingüino(Tweety)$ *al programa, entonces el programa tendrá ahora la posibilidad de construir un argumento para* $\neg vuela(Tweety)$ *que derrota al argumento para* $vuela(Tweety)$, *es decir, el programa se protege de la observación anómala.*

EJEMPLO **5** *Siguiendo nuevamente con el Ejemplo 1 desde el programa* P_1. *Tenemos en* $vuela(Schmuck)$ *una laguna. En esta situación es necesario o bien establecer un ranking entre los condicionales derrotables utilizados para generar los argumentos para* $vuela(Schmuck)$ *y para* $\neg vuela(Schmuck)$, *o bien es necesario modificar alguno de los condicionales (por ejemplo,* $pingüino(X) \wedge en\text{-}avión(X) \succ\!\!\!- vuela(X)$).

Referencias

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

[2] Claudio Delrieux. The Rôle of Defeasible Reasoning in the Modelling of Scientific Research Programmes. En *IC-AI 2001 Conference*, págs. 861–868, ISBN 1-892512-81-5, 2001.

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

[4] Imre Lakatos. *Proofs and Refutations. The Logic of Mathematical Discovery.* Cambridge, 1976.

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

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

[1] Es decir, la explicación que realiza la menor cantidad de suposiciónes.