Abducción, Razonamiento Revisable y Explicación Científica

Claudio Delrieux
Departamento de Ingeniería Eléctrica
Universidad Nacional del Sur - Alem 1253 - (8000) Bahía Blanca
claudio@acm.org

1 Razonamiento revisable

La necesidad de contar con modelos de razonamiento no monotónico en los sistemas inteligentes fue rápidamente reconocida en la comunidad del KR&R [3], especialmente para poder manejar los problemas originados al aplicar deducción en teorías incompletas. Las reglas derrotables en particular, y la inferencia ampliativa en general, pueden considerarse como un mecanismo heurístico para *construir* nuevo conocimiento a partir del conocimiento ya disponible. Este conocimiento generado es de naturaleza tentativa, es decir, es aceptable en tanto que no se modifique el contexto dentro del cual fue producido.

En el razonamiento revisable surgido a partir de la década del 80 se buscó la manera de respetar la *forma* de la deducción por medio de esquemas de inferencia similares al *modus ponens*. De esa manera, representamos con una premisa *'Normalmente, cuando a sucede, entonces b también sucede'* la predisposición a realizar la inferencia no monotónica de b cuando en el contexto de la teoría es posible demostrar a pero no es posible demostrar $\neg b$. Dicha premisa asume la forma de una regla y por lo tanto se suele denominar *regla revisable*. Los sistemas de razonamiento no monotónico surgidos recientemente buscan solucionar el problema del *encadenamiento* de una línea de razonamiento, es decir, la construcción de una secuencia análoga a una demostración para una conclusión. Estas demostraciones con reglas revisables suelen denominarse *argumentos* o *teorías*.

2 Abducción con reglas revisables

Como fuera mencionado en [2], es posible utilizar otros esquemas ampliativos (sintéticos) de inferencia dentro del marco formal del razonamiento revisable. En particular, en este trabajo deseamos enfocarnos específicamente en la abducción. La abducción es una forma de razonamiento que ha sido caracterizada como la inferencia de la mejor explicación para un hecho o átomo de conocimiento particular. Un hecho es explicado si puede deducirse de la teoría, posiblemente luego de asumir ciertas hipótesis explicativas. Si no hubo hipótesis, el hecho fue simplemente deducido. Si se requirieron hipótesis, estas fueron inferidas por abducción, y si hubiese sido necesario agregar una nueva regla a la teoría, en ese caso suele mencionarse la palabra *inducción*.

No vamos a mencionar aquí la importancia de la abducción en general, y en la IA en particular (ver [11]). Sin embargo, es importante destacar que su tratamiento ha sido en general insatisfactorio, es decir, se considera que la abducción debe ser tratada como una deducción más ciertos procedimientos. En [6], por ejemplo, la abducción en teorías monotónicas es manejada deductivamente por medio de la clausura del dominio junto con requisitos de consistencia. En efecto, en teorías cerradas se puede asimilar la abducción directamente a la deducción en una teoría transformada que se obtiene a partir de la teoría original [1]. Es más, sus propiedades computacionales son similares a las de la inferencia no monotónica, es decir, es en general intratable, aunque se han identificado casos interesantes que son computacionalmente tratables [4].

Sin embargo, un aspecto de gran importancia para el estudio de la abducción surge de su

comportamiento frente a la existencia de reglas revisables en la teoría. En este caso resulta evidente que no puede asimilarse la abducción a una teoría transformada. Es más, como veremos en la siguiente Sección, surgen varias asimetrías entre el uso "deductivo" de una regla revisable[1] y su uso "abductivo".

Utilizaremos en este trabajo una definición de demostración derrotable $|\sim$ basada en los sistemas argumentativos abstractos de Vreeswijk [14]. Las reglas de inferencia son o bien estrictas (similares a la implicación material) o bien derrotables, donde el formato general es

$$s_1, s_2, \ldots, s_n \rightarrow t,$$

siendo cada elemento del antecedente y el consecuente un miembro del lenguaje L.

Para representar la inferencia abductiva lo ideal sería poder utilizar un esqema de inferencia

$$\frac{b(t) \quad |\sim_T b(t) \quad a(X) |\sim_T b(X)}{a(t),}$$

es decir, *si se produce la observación de un hecho sorpresivo* b(t) *el cual no es predicho por la teoría revisable* T, *pero encontramos que* a(X), *agregado como antecedente a la teoría, permite generar una demostración derrotable para* b(X), *entonces inferimos* a(t).

El encadenamiento de reglas origina los argumentos. En Vreeswijk un argumento (básico) es un elemento de L, o bien una fórmula

$$a_1, a_2, \ldots, a_n \rightarrow b,$$

donde cada miembro del antecedente es un (sub)argumento, y existe una regla

$$s_1, s_2, \ldots, s_n \rightarrow b,$$

tal que el consecuente de a_1 es s_1, etc. con la restricción de que b no aparece en ningún subargumento.

De esa forma, un esquema de inferencia deductiva queda representado por un argumento de la forma

$$b(t), (a(X) \rightarrow b(X)) \rightarrow a(t),$$

donde el requisito de no demostrabilidad previa de b(t) es innecesario dada la restricción de circularidad vista más arriba. Como se puede apreciar en esta representación, las abducciones podrían encadenarse hacia atrás. Esto da lugar a especular de qué manera debe agregarse al sistema la posibilidad de incluir derrotadores entre argumentos abductivos, por ejemplo, a través de mecanismos al estilo de la especificidad como los propuestos por [7, 10], el acopio de razones [12,8], o la comparación del poder predictivo o explicativo de las hipótesis abducidas. En este trabajo comentaremos solamente este último aspecto, porque es donde parecen darse las situaciones que mejor distinguen al razonamiento abductivo de otros tipos de razonamiento.

3 Expicación científica

Como mencionáramos más arriba, existen asimetrías entre el uso deductivo y el uso abductivo de una regla revisable. En la teoría de la ciencia, el primer caso suele denominarse *predicción*, es decir, cuando los hechos relevantes y la teoría están establecidos y permiten predecir o retrodecir un

determinado fenómeno. El segundo caso, el uso abductivo, en general se asocia al contexto de descubrimiento en el uso de la teoría, denominándose *explicación* [5].

Paradójicamente, las explicaciones con reglas revisables normalmente son más fuertes que las predicciones. Decimos paradójicamente, porque con una regla estricta (una implicación material), el uso deductivo es correcto (sound) y por lo tanto el razonamiento es demostrativo, mientras que el uso abductivo es incorrecto (unsound) y por lo tanto la conclusión no tiene valor demostrativo. Sin embargo, en una regla revisable las cosas parecen suceder de otro modo.

Consideremos la regla *tomar mate con un engripado puede producir contagio*. Afirmar que X se va a contagiar la gripe por tomar mate con su esposa enferma es una predicción de relativo poder. Sin embargo, si X contrajo gripe, y luego refexionando recuerda haber tomado mate con un compañero de oficina que se sentía mal, la misma regla, utilizada como explicación, es fuerte.

Otra asimetría notable ocurre cuando se considera la posibilidad de hacer *acopio* de razones (*accrual of reasons*). El acopio de razones consiste en reforzar el poder demostrativo de una conclusión si ésta es soportada por dos o más líneas independientes de razonamiento. Este principio fue apenas defendido por pocos autores [8, 12], aunque luego fue casi universalmente abandonado [9, 13]. Un ejemplo práctico para entender por qué el acopio de razones es insostenible consiste en considerar el llamado "efecto invernadero". Existen tres causas conocidass que relacionan la densidad de anhidrido carbónico con la temperatura en la atmósfera. Dos de estas causas hacen disminuir la temperatura si la densidad aumenta, mientras que la tercera opera a la inversa, siendo sin embargo dominador en el efecto global.

En el caso de la inferencia abductiva, este principio parece natural, y de hecho se utiliza normalmente en el razonamiento científico. Por ejemplo, una de las explicaciones cosmológicas más famosas para dar cuenta de la deriva hacia al rojo de los espectros luminosos de galaxias lejanas es la teoría del *big bang*, por la cual el universo se creó en un momento determinado y desde entonces los grandes grupos de galaxias se alejan entre sí a velocidad uniforme. Si bien esta teoría tiene un determinado poder explicativo, lo más interesante es que el mismo aumenta si se consideran otras consecuencias, por ejemplo la recientemente observada radiación de fondo[2].

Pensemos en la siguiente situación:

$$a \rightarrow c,$$
$$b \rightarrow c,$$
$$b \rightarrow d.$$

¿Qué explicaciones podemos encontrar para justificar el hecho observado c? Por inferencia abductiva podemos encontrar que tanto a como b son explicaciones posibles. ¿Cuál de dichas explicaciones parece la más plausible? Si el hecho d no es observado, entonces la explicación b pierde credibilidad y tal vez a es preferible como explicación, mientras que si d es observado, entonces se da una situación inversa, porque b se ve "más confirmada" que a.

Notas:

[1]Abusando del lenguaje, denominamos de esta manera la aplicación de la regla utilizando un esquema "hacia adelante" similar al *modus ponens*, para contraponerlo con un uso "hacia atrás" en un esquema similar a la abducción.

[2]En realidad la existencia de la radiación de fondo fue calculada y predicha como consecuencia del *big bang* mucho antes de ser observada. Pero esto es solamente una circunstancia histórica. Podría haberse dado el caso de que se observara la radiación de fondo y se predijera la deriva espectral. Lo importante de nuestro punto aquí es que si ambos fenómenos fueran observados sin teoría subyacente, entonces una teoría que explique ambos fenómenos tiene mayor poder explicativo que cualquier teoría que explique a uno solo de ellos.

Referencias

[1] Luca Console, Luigi Portinale, y Daniele Theseider. On the Relationship Between Abduction and Deduction. *Journal of Logic and Computation*, 1(5):661-690, 1991.

[2] Claudio Delrieux. Inferencia Ampliativa y Razonamiento no Monotónico. En *WICC 99*, páginas 75-79, San Juan, 1999.

[3] Matthew L. Ginsberg (ed.). *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann Publishers, Los Altos, California, 1987.

[4] T. Eiter y G. Gottlob. On the Complexity of Propositional Knowledge Base Revision, Updates and Counterfactuals. *Artificial Intelligence*, 57(2-3):227-270, 1992.

[5] Carl G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press, New York, 1965.

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

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

[8] John Pollock. *Cognitive Carpentry*. MIT Press, Minneapolis, 1995.

[9] John Pollock. Degrees of Justification. En P. Weingartner, G. Schurz, y G. Dorn, editores, *The Role of Pragmatics in Contemporary Philosophy*, páginas 207-223. Verlag Hölder-Pichler-Tempsky, Wien, Österreich, 1998.

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

[11] Pietro Torasso, Luca Console, Luigi Portinale, y Daniele Theseider. On the Role of Abduction. *ACM Computing Surveys*, 27(3):353-355, 1995.

[12] Bart Verheij. Two Approaches to Dialectical Argumentation: Admisible Sets and Argumentation Stages. In *Proceedings of the Eight Dutch Conference on Artificial Intelligence*, páginas 357-368, Utrecht, Netherlands, 1996. NAIC'96, Universiteit Utrecht.

[13]Bart Verheij. Argue!, an Implemented System for Computer-Mediated Defeasible Argumentation. En *Proceedings of the Tentht Netherlands/Belgium Conference on Artificial Intelligence*, páginas 57-66, Netherlands, 1998. NAIC'98, Amsterdam.

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