La difusa frontera entre biología y computación (o hacia la mecanización de la ciencia)

RODERIC GUIGÓ I SERRA

Si introducimos en Google la palabra «Bioinformatics» —el término que denota en inglés la disciplina científica en la cual la Informática es utilizada en la investigación en Biología—, la búsqueda subsiguiente en Internet nos proporciona más de 100 millones de documentos. Un número que no es muy distinto del número de documentos en Internet que contienen palabras como Genética, Bioquímica, Fisiología, Ecología, etc.; todas ellas términos que denotan disciplinas de la Biología con una historia centenaria. La Bioinformática, sin embargo, es una disciplina novísima: Medline —la base de datos que compila la literatura científica en Biología y Medicina— no incluye ningún artículo en el que aparezca el término «Bioinformatics» con anterioridad al año 1990. En estos momentos, sin embargo, son cerca de 12.000, de los cuales 2.500 han sido publicados durante el último año. Hemos sido, pues, testigos, en el corto espacio de una década, de la explosión de una nueva disciplina científica —una explosión insólita en la historia de la ciencia. Se atribuye al desarrollo de las tecnologías de la Genómica, el papel clave que la Bioinformática desempeña hoy en día en la investigación en Biología. En efecto, las tecnologías de la Genómica generan un volumen de datos, sobre los fenómenos de la vida, de una magnitud sin precedentes en Biología. Los experimentos con matrices de ADN (los «DNA microarrays» en inglés) son ilustrativos en este sentido. Mediante estas matrices, cuyo soporte físico apenas supera unos pocos centímetros cuadrados, es posible obtener en pocas horas datos sobre el comportamiento simultáneo de miles de genes bajo unas condiciones determinadas. Miles de experimentos con matrices de ADN se llevan a cabo diariamente en todo el mundo, muchos de ellos de forma casi automática. Hace sólo una década, sin embargo, la obtención de esos mismos datos sobre un único gen era el resultado del trabajo continuado de un investigador (o de un equipo de investigadores) durante meses, a menudo durante años. La Biología ha pasado, pues, en muy poco tiempo, de ser una ciencia en la cual el esfuerzo humano se orientaba principalmente hacia la obtención de (unos pocos) datos, a ser una ciencia que genera un volumen literalmente vertiginoso de datos sin apenas esfuerzo humano. En este nuevo paisaje en el que se desarrolla la investigación en Biología —en Genómica, para ser más precisos—, los métodos informáticos desempeñan, pues, un papel esencial, tanto en la planificación de los experimentos, como en su ejecución, como, sobre todo, en el almacenamiento y análisis de sus resultados; es decir, en el proceso mediante el cual los datos experimentales son convertidos en «conocimiento»: en la formulación de hipótesis sobre las relaciones de causalidad entre fenómenos naturales. Hipótesis que, en el relato habitual de la práctica científica, dan lugar, a su vez, a la planificación de nuevos experimentos. Aunque aceptamos la importancia de la contribución de la computación, consideramos en general que este proceso de reflexión dialéctico —de interrogación de la naturaleza mediante la experimentación y de interpretación de los datos resultantes mediante hipótesis explicativas— es, en sus componentes esenciales, un privilegio del

ser humano, y que la actividad científica es, en consecuencia, una actividad inseparable de la humanidad. En Genómica, sin embargo, el paradigma empieza a tambalearse. La creciente automatización y robotización la práctica experimental hace que la producción de datos crezca a un ritmo exponencial. Los ordenadores son también, es cierto, cada vez más potentes y los programas informáticos de análisis cada vez más sofisticados. Pero sin el proceso de reflexión intelectual que llevan a cabo los científicos, los datos no se convierten en conocimiento. Y el número de científicos no crece exponencialmente. El resultado: millones de datos, obtenidos mayoritariamente por robots de forma automática, se acumulan en sistemas informáticos de almacenamiento, a la espera de que algún científico los convierta en conocimiento. Pero éste, cada vez con más frecuencia, nunca llega y el proceso científico se interrumpe; datos (secuencias genómicas de organismos remotos, valores de expresión de miles de genes en centenares de condiciones experimentales distintas, estructuras tridimensionales de centenares de proteínas...) que no generan hipótesis, ni, en consecuencia, dan lugar a la planificación de nuevos experimentos. Ante esta situación, una de las soluciones que ha empezado a ensayarse en Bioinformática consiste en la ampliación del espacio de las máquinas en la investigación científica: en consignar también a las máquinas tareas del método científico que habíamos considerado hasta ahora intrínsecamente humanas. En efecto, en el año 2004, Ross King y sus colaboradores crearon el primer científico robot (King *et al.*, 2004). Este robot no sólo llevaba a cabo experimentos —una tarea que los robots llevan a cabo comúnmente en muchos laboratorios de Biología— sino que interpreta los resultados y genera nuevas hipótesis, a partir de las cuales plantea, a su vez, nuevos experimentos. El robot fue inicialmente diseñado para resolver una tarea típica en Génomica: elucidar la función de un conjunto de genes en la levadura. Y parece ser que llevó a cabo esa tarea de manera más eficiente que un estudiante de doctorado; ambos, robot y estudiante, llegaron a las mismas conclusiones, pero para llegar a ellas el robot planteó un número menor de hipótesis y, en consecuencia, realizó un número menor de experimentos. Con el robot científico, en efecto, trasladamos a las máquinas actividades consideradas intrínsecamente humanas: la interpretación, la reflexión, la planificación...; se mecaniza el pensamiento, como habría intuido hace siglos Ramon Llull. Asistimos pues, precisamente en el campo de la Bioinformática, a los primeros intentos de automatizar la investigación científica y autonomizarla del ser humano.

Puede argumentarse que la generación masiva de datos no es patrimonio de la investigación en Biología, y que el mismo fenómeno tiene lugar en áreas tan diversas de la ciencia como la astronomía, la economía, la física de altas energías, la meteorología etc. Y que, en todas ellas, los métodos computacionales desempeñan un papel esencial en el tratamiento e interpretación de los datos. Es cierto. ¿Por qué, pues, el científico robot surge en el campo de la Bioinformática? ¿Por qué, como podemos comprobar de nuevo en Google, existen apenas documentos en Internet en los que aparezcan términos como «astroinformatics», «meteoinformatics», «econoinformatics», en contraste con las decenas de millones de documentos en los que aparece el término «Bioinformatics»? Me atrevo a aventurar una hipótesis: La relación entre Biología y Computación se establece a otro nivel, es más íntima que la que existe entre la Computación y otras disciplinas científicas. Quizás por razones históricas —el desarrollo de la Biología Molecular discurre paralelo al desarrollo de la Informática; muchos de los términos del vocabulario básico de la Biología Molecular, que se establece durante los años sesenta, parecen originados en la Informática: código o programa genético, traducción, transcripción, procesamiento, etc.—, el hecho es que la metáfora computacional forma parte del discurso que vertebra la Biología Molecular. Discurso de acuerdo con el cual los organismos vivos son el resultado de la computación de la secuencia de su genoma. ¿Pero se

trata realmente de una metáfora? ¿Es que podemos imaginar ejemplos más nítidos de lo que constituye una computación que los procesos básicos y universales de la vida: la Replicación del ADN, la Transcripción del ADN a ARN y la Traducción del ARN mensajero a proteína? Tomemos el caso de la Traducción, el proceso cuyo input es la secuencia de nucleótidos de un ARN mensajero (una secuencia en un alfabeto de cuatro letras: ACGT), y cuyo output es la secuencia de aminoácidos de una proteína (una secuencia en un alfabeto de 20 letras: ACDEFGHIKLMNPQRSTUWY). Las operaciones mediante las cuales el input dé lugar al output se encuentran especificadas, en este caso, en el Código Genético, la tabla de equivalencias que asigna a cada una de las 64 posibles combinaciones de tres letras en el alfabeto de los nucleótidos una letra en el alfabeto de los aminoácidos. ¿Existe un ejemplo mejor de computación que la Traducción?

En la célula viva, los mecanismos mediante los cuales se llevan a cabo estas operaciones se fundamentan en la capacidad de hibridación que exhiben moléculas de ácidos nucleicos complementarias —la cual permite el reconocimiento específico de secuencias, es decir, en términos informáticos, su direccionamiento— y en la capacidad de edición de las secuencias por parte de enzimas de restricción y otras proteínas. Propiedades ambas —direccionamiento y edición— en las que, por otra parte, se fundamentan también muchas estrategias experimentales en Biología Molecular —en particular, los experimentos con matrices de ADN— y que las hace, como los procesos básicos de la vida, conceptualmente indistinguibles de la ejecución de un programa. En un caso se trata de operaciones sobre una secuencia binaria de 0s y 1s, en el otro, sobre una secuencia de As, Cs, Gs y Ts. Adleman intuyó hace más de 10 años esta equivalencia conceptual y en su artículo seminal «Molecular Computation of Solutions to Combinatorial Problems» (Adleman, 1994), describía la solución, mediante la utilización de secuencias de ADN, del conocido problema en Computación consistente en determinar si un grafo es hamiltoniano; es decir, si existe un camino que recorra todos los nodos del grafo sin pasar por un mismo nodo más de una vez. Adleman codificó una pequeña versión de este problema —un grafo de 7 nodos— no en la secuencia binaria de los ordenadores digitales, sino en la secuencia de nucleótidos de moléculas de ADN y determinó su solución mediante un algoritmo que ejecutó en tubo de ensayo, utilizando técnicas estándar de Biología Molecular. Al demostrar que el ADN tiene capacidad de calcular, Adleman inaugura la Computación Molecular, y al mismo tiempo contribuye a diluir la frontera entre vida y computación.

Bibliografía

ADLEMAN, L.M. (1994), «Molecular computation of solutions to combinatorial problems», *Science*, 266: 1.021-1.024.

KING, R.D., WHELAN, K.F., JONES, F.M., REISER, P.G., BRYANT, C.H., MUGGLETON, S.H., KELL, D.B., OLIVER, S.G. (2004), «Functional genomic hypothesis generation and experimentation by a robot scientist», *Nature*, 15: 247-252.