26/09/2016

Bioinformatics 101

TXT LIONEL URAN LANDABURU IMG FELIPE NIÑO

¿Qué es la bioinformática?

Para los de veintipico, como yo, la compu fue un pivote fundamental de nuestra infancia. Mis viejos la habían comprado allá por el '97 para montar un sistema informático para el negocio. Windows 95... Abrir y cerrar carpetas, crearlas y eliminarlas, vaciar la papelera, hacer dibujitos en paint e imprimirlos; todo era una fiesta a mis 7-8 años, y no tardamos en ser amigos. La compu y yo, un solo corazón. Como yo, en aquel entonces un montón de otros privilegiados crecimos al candor de una Pentium I bien ruidosa y llena de amor. Hoy, poco menos de 20 años más tarde, todo pasa por una compu (cualquiera sea la que tengas en el escritorio o en el bolsillo) e internet. En Argentina, el último censo oficial del INDEC (sí, 2015, y no, no vamos a entrar en discusión sobre papeles más o menos flojos) señaló que el

acceso a distintos formatos de Tecnologías de la Información y Comunicación aumentó en categoría televisores (+0.2 %), teléfonos móviles (+2.9 %), computadoras (+10.6 %) e internet (+13.8 %). Más allá de cualquier discusión que podría bullir, este aumento es consistente con una tendencia mundial. Más personas tenemos compu e internet, y eso es dato.

Obvio que los laboratorios no van a estar exentos de esa tendencia (lejos de exentos, recordemos dónde nace Internet). Hoy en día, en el laboratorio, contar con un par de computadoras es prácticamente indispensable para leer, analizar y entender los datos que producimos a través de la experimentación. Ahora, hay dos formas de ver esto: una es entender la compu como un re buen cuaderno de laboratorio; uno que nos permite tomar nuestros datos y presentarlos de manera que podamos recorrerlos, revisarlos, reverlos y, recién ahí, usar nuestras cabezas para encontrar patrones y hacer inferencias nuevas. Hasta ahí, la computadora es un GRAN cuaderno, pero un cuaderno al fin.

Ahora, ¿qué pasa si queremos más? O, peor, si lo necesitamos. ¿Qué pasa cuando la cantidad de información generada es tanta que escapa todo intento humano de racionalización? Este límite sobre nuestra capacidad humana de interactuar con información generada en sistemas digitales es lo que me atrevo a denominar 'Síndrome de Vacaciones en Las Toninas'. El mismo refiere al momento exacto de la historia humana donde por primera vez logramos que la capacidad de capturar y guardar información nos superase tan evidentemente como para preguntarnos cuándo fue la última vez que vimos las 378 fotos de vacaciones que sacamos usando nuestra primera cámara digital con memoria SD.

Evidentemente, nos enfrentamos al límite humano para la contemplación de imágenes de sombrillas, churros, olas y viento. Ese mismo límite, por suerte, lo enfrentamos los científicos gracias a nuestra capacidad actual de adquirir información del entorno y digitalizarla. Bueno, las fotos de vacaciones ya están. ¿Y ahora? ¿Cuáles de las 9735589 miro? Thor bendiga la bioinformática.

Tengo fotos de vacaciones. Igual, imprimí las mejores para verlas

Lejos de ser el control de organismos mitad robot, mitad dinosaurio gigante, mitad mejormascotadetodoslostiempos mediante métodos electrónicos, la bioinformática es *la investigación, desarrollo o aplicación de herramientas computacionales y aproximaciones para la expansión del uso de datos biológicos, médicos, conductuales o de salud, incluyendo aquellas herramientas que sirvan para adquirir, almacenar, organizar, analizar o visualizar tales datos*. Definición aburridísima que podemos resumir en 'usar computadoras para abarajar información biológica que ni a palos podríamos, por más papel, lápiz, tiempo y cafeína que le tiráramos al problema'. En el fondo, lo que queremos es asistirnos, programando alguna forma de separar paja y trigo, especialmente cuando nos enfrentamos a marañas de datos tan grandes que son o tediosas o directamente imposibles de manejar.

En esta necesidad de buscar la mejor forma de masticar datos e identificar patrones, nos encontramos con ese vértigo que da saber que, para progresar, necesitás tocarles el timbre a disciplinas vecinas. Así es que un biólogo cargado de datos recorre el camino que hay entre su pabellón y el de matemática, física y computación mientras mastica la idea de que asumir compartimentos estancos en las áreas del conocimiento es un embole. Yo te voy a dar estos datos que saqué de un sistema biológico y convertí en digital, nos vamos a sentar juntos para tratar de entender qué patrones identificar, vos vas a hacer tu cosita de amase de información y, con las agujas que encontremos en el pajar de datos (porque si

vamos con paja y trigo, vamos con aguja, ya fue todo, la muerte de la metáfora), yo trataré de entender algo que antes me era invisible en mi sistema biológico inicial solamente por estar oculto entre otra pila de factores.

Es a través de la bioinformática que hoy podemos saber qué tan lejos, evolutivamente, estamos los humanos de las ratas; o qué secuencias en el ADN conllevan un riesgo intrínseco de padecer ciertas enfermedades. **Podemos usar la bioinformática para proponer herramientas de diagnóstico, para buscar nuevas drogas para algunas enfermedades o encontrar nuevos *targets* para atacar organismos patógenos.**

El nexo clave entre uno y otro es la **informatización de la biología**, es decir: expresar la biología en un formato interpretable por una computadora. Esta idea tiene un giro histórico clave con el nacimiento de la **biología molecular**. Es ella la que ha sido capaz de darle identidad informática a proteínas y ácidos nucleicos, a través de la expresión de estas **macromoléculas** como la sucesión de un montón de letritas que representan sus **secuencias**. Cada letra en una secuencia representa un eslabón en la cadena (nucleótidos en el ADN, aminoácidos en las proteínas). El núcleo mismo de la vida entendida como información.

Hecha esta elipsis, podemos volver al Síndrome de Vacaciones en Las Toninas y a ese álbum tuyo que tiene 500+ fotos en el que, fácil, 100 son de los sucesivos fracasos en el intento de saltar todos al mismo tiempo. Cuando tenías esa cámara de rollo donde no sabías cómo habías salido y revelar te costaba una pila de billetes, eras una persona de mesura, un ícono del control, un paladín de la administración de recursos. Pero bueh, llegaron las cámaras digitales y ¿¿EN QUÉ TE HAS CONVERTIDO?? Generadores compulsivos de fotografías. Derrochadores de bytes. Ilusos, aferrados al afán de que en alguna foto hayamos salido bien todos.

Con la bioinformática pasa un poco lo mismo. Al principio de los tiempos (de los tiempos bioinformáticos, claro, principios del ’80), la comunidad científica contaba con tan sólo unas cuantas secuencias (algo así como 600). Si la idea era estudiar un gen o una proteína en particular, capaz sólo había una o dos secuencias. Con suerte. Mucha. Puesto así, poco podía ofrecer la informática para resolver los problemas biológicos de antaño. Salvo que venga una recontra

revolución de la secuenciación de ADN y proteínas y de golpe tengamos un montón de montones de pilas de información gracias a la disminución de costo y el tiempo de respuesta. En particular, para ácidos nucleicos, la aparición de nuevas tecnologías de secuenciación propone la posibilidad de elegir y, con ello, un escenario de competencia (como en cualquier mercado) en el que el precio podría resultar crítico.

Para ponerlo en números tangibles, el Proyecto Genoma Humano se extendió por 13 años y tuvo un costo aproximado de 2500 millones de dólares + expensas + 2 meses de depósito. Hoy en día se puede secuenciar el genoma de cualquiera de nosotros por poco más de 1000 dólares y en cuestión de días. Y acá estamos, cada uno con su álbum de las Toninas y su cámara de fotos secuenciadora, generando fotos de genomas porque se puede y es más o menos barato.

Secuencias de mamífero marino

Esto supone un **desafío informático** grande, y una **oportunidad biológica** gigante. Es como cuando vas a la feria del libro el día de liquidaciones y te comprás más libros de los que vas a poder leer. ¿De nuevo la analogía del libro? Sí. ¿Podría usar otra? Obvio. Pero la del libro, además de ser bastante conceptualmente adecuada, es re linda, porque los libros son re lindos.

Leer todos esos libros te llevaría mucho tiempo, pero ¿no vale la pena? ¿Y si pudieras hacer que una computadora leyera todos los libros y te contara más o menos de qué va lo importante? Sacando la fantástica dimensión experiencial de leer propiamente un libro; si la idea es adquirir la información que está inmersa entre las letras, ¿no lo harías? Una onda Matrix pero sin la punción lumbar.

Esa es una de las cosas que hacemos los bioinformáticos: le pedimos a la compu que lea genomas enteros, repetidos 1000 veces o secuencias proteicas de a montones. Le pedimos que lea el mismo libro, impreso por distintas editoriales, muchas veces. Y le decimos que, cuando lo haya hecho, nos cuente algo. A veces le pedimos que nos cuente si halló diferencias entre secuencias 'iguales' (los mismos genes), comparando variantes (estrictamente, alelos) de personas con alguna condición patológica y personas sin esa patología, para entender si la patología tiene un correlato genético observable. Otras veces le pedimos que encuentre una región genética altamente conservada (es decir, que sea igual en todos) entre especies relacionadas. **En términos evolutivos, si algo es importante, que mute es muy negativo, y esos bichos palman, entonces es la misma selección natural la que se encarga de que esa región aparezca conservada.** Esto nos sirve no solamente para encontrar parentescos entre especies, sino para cosas como saber dónde atacar un organismo patógeno. Cuando tenemos suficiente cafeína en sangre, nos animamos a preguntarle si le parece viable (energética y espacialmente) que cierta droga interaccione con una proteína para poder proponer, con cierto grado de certeza, que esa droga podría resultar en un tratamiento viable.

Claro que buena parte de nuestra ciencia tiene sólo un carácter predictivo y todos nuestros hallazgos son propuestas que habrán de ser comprobadas experimentalmente más tarde. Armamos modelos que arman modelos que después probamos; de esas pruebas, aprendemos y los mejoramos. Como todos los científicos.

Una aproximación así sobre una pregunta biológica tiene la ventaja de ofrecer una capacidad de prueba ridículamente superior a la del laboratorio experimental. Por ejemplo, a un técnico en el laboratorio probar 1000 drogas podría llevarle años y el costo experimental sería altísimo. Un bioinformático, por su parte, tiene la capacidad de probar 1000 drogas en cuestión de semanas, seleccionar las 10 más aptas y ofrecer esa información a la comunidad científica para que las otras 990 sean desestimadas.

Otra función no tan cool (pero no por eso menos importante) de la bioinformática es la de desarrollar repositorios y bases de datos capaces de guardar toda la

información que genera la comunidad científica, para que sea accesible a todos y podamos usarla lo más eficientemente posible. Qué lindo pensar en escalar eso, tener un repositorio internacional, abierto, bien curado, con los experimentos registrados previamente, los datos abiertos, los crudos de todo. Es mi tierra sin abogados de Lionel Hutz.

En algún momento, las computadoras se metieron en todo, también en los laboratorios. Las usamos para hablarnos entre nosotros, para compartir información, para guardarla, pero en algún momento fuimos por más. En algún momento tuvimos tantas cosas para hacer, tantas preguntas nuevas, tantos factores a comparar al mismo tiempo y tantas posibilidades técnicas para tratar de hacer algo con esos datos que el mundo físico nos limitó y terminamos invirtiendo la lógica: hicimos laboratorios que viven adentro de computadoras y los fuimos habitando. Somos los ecólogos de rodillas limpias, los genetistas secos, los moleculares que no pipetean. Los que pavimentamos el *in silico*.

Referencias

INDEC (Instituto Nacional de Estadísticas y Censos) (2015), "**Encuesta Nacional sobre Acceso y Uso de Tecnologías de la Información y la Comunicación (ENTIC)**" [Informe preliminar sobre indicadores básicos de acceso y uso. Resultados de mayo-julio de 2015]

Sumate en
eglc.ar/bancar