

BioIA la evolución de la Bioinformática

Introducción

Hay una rama dentro de las ciencias biológicas llamado biología computacional, que utiliza métodos computacionales para resolver problemas de la biología, y la que a su vez contempla internamente algunas subáreas como: biología molecular computacional, biología computacional de sistemas, bioinformática, entre las principales. Los avances en computación sin duda han ido en directo apoyo a estas áreas, desde que Alan Turing y el grupo de Los Alamos, en los años 50 implementara un modelo de morfogénesis biológica (desarrollo de patrones y formas en organismo vivos) pasando por modelamiento de códigos genéticos, la comprensión de las estructuras de las proteínas, incluso determinar su estructura tridimensional.

Desarrollo

La rápida evolución de la computación en el siglo XX permitió evolucionar desde el estudio de las secuencias y estructuras de moléculas biológicas al análisis de su funcionamiento. Predecir el funcionamiento implica evaluar la secuencia y su estructura y analizar las interacciones entre una proteína y otras moléculas, lo que resulta en utilizar una gran capacidad de computo y herramientas computacionales sofisticadas, como HPC (computación de alto rendimiento), esta potencia abrió una extensión de la experimentación tradicional, basada en hipótesis hacia el análisis basado en datos, que permite realizar experimentos computacionales o también llamado in silico, a escalas sin precedentes.

Hitos importantes como el dogma central de la Biología Molecular propuesto en 1956, y el Proyecto Genoma Humano, desarrollado entre 1990 a 2003, que entrego información sobre 3 mil millones de pares de bases químicas (A,T,G,C) que componen el ADN humano, y 20 mil a 25 mil genes, aquí es donde la bioinformática a sido clave y demuestra todo su poder para analizar los datos antes mencionados. Hoy en día se utilizan bancos de datos que organizan y almacenan esta información, entre las mas importantes esta EMBL-Bank en Reino Unido, DDBJ en Japón, GenBank de NCBI de Estados Unidos.

Luego del descubrimiento de la estructura de la doble hélice del ADN por James D. Watson y Francis H. C. Crick en 1953, estas técnicas de procesamiento a través de software se extienden a secuencias de ADN y aminoácidos, con métodos computacionales que utilizaban complejos algoritmos y programación dinámica, dada la escala de datos disponibles, que incluyen algoritmos de grafos, agrupamiento, búsqueda de patrones.

La IA (inteligencia artificial), sobre todo los algoritmos de ML (machine learning) y DL (deep learning) aplicado a los campos de genómica, proteómica y metabolómica, ha revolucionado el análisis de los complejos dataset, ayudando a los científicos a entender nuevos procesos biológicos, diagnósticos de enfermedades y diseño de terapias personalizadas, acá juegan un papel crucial las CNN (redes neuronales convolucionales) y RNN (redes neuronales recurrentes) que han ayudado a encontrar funciones diagnosticas, pronosticas y terapéuticas, por ejemplo, en pacientes con Cáncer de mama.

También algoritmos de GNN (redes neuronales graficas) y GAN (redes neuronales adversariales) ayudan a predecir cómo se comportan potenciales moléculas incluyendo la predicción de estructuras de proteínas y clasificación de enzimas que son cruciales para el desarrollo de nuevos fármacos.

Conclusión

La IA ha alcanzado innumerables áreas de la ciencia, que se benefician de sus algoritmos, acá se mostró brevemente como es capaz de codificar el conocimiento biológico y procesar incluso en lenguajes natural, desde modelado de secuencias para reconocimiento de patrones al análisis de datos complejos para estudios de expresión génica a gran escala, hasta para extraer información de la literatura científica, donde ha sido un gran aporte los algoritmos de aprendizaje automático.

Christian Parra Villena

Presidente del Consejo de Especialidad de Computación e Informática

Colegio de Ingenieros de Chile