

INVESTIGADOR MEXICANO ENSAYA EN LAS LIGAS MAYORES DE LA COMPUTACIÓN MODELOS PROBABILÍSTICOS CON APLICACIONES EN BIOLOGÍA

José Luis Carrillo Aguado

Periodista científico del IPN.

El doctor Manuel Reyes Gómez, investigador de Microsoft Research (la división de investigación de Microsoft Corporation) y sus colegas trabajan en un proyecto cuyo objetivo es el desarrollo de modelos probabilísticos que caractericen los diferentes procesos del sistema inmunológico. El modelo más sobresaliente desarrollado por este equipo permite la estimación de la energía de enlace necesaria para la unión de proteínas. Las proteínas son secuencias de aminoácidos, las cuales se doblan y desdoblan, formando estructuras tridimensionales. Las proteínas regulan la mayoría de las funciones de la célula.

Esta investigación puede resultar útil en el estudio de los patrones de mutación de virus como el de la inmunodeficiencia humana (VIH). El objetivo es analizar el virus del SIDA, para contribuir al entendimiento del comportamiento de este ser –que está entre lo vivo y lo inerte– y estar así en posibilidades de desarrollar una vacuna contra dicho virus.

Para desarrollar su proyecto, el doctor Manuel Reyes Gómez tomó en cuenta los diferentes tipos de interacciones moleculares. El modelo propuesto por su equipo aprende de manera automática a caracterizar los distintos tipos de interacciones moleculares, como los enlaces covalentes, los enlaces iónicos, los enlaces de hidrógeno y las atracciones de Van Der Waals. El modelo toma en cuenta dos tipos de parámetros: uno de ellos mide la importancia de la interacción entre dos moléculas de aminoácidos dada su posición en la secuencia de aminoácidos, independientemente del tipo de aminoácido; otro grupo de parámetros estima la importancia del tipo de aminoácido en dichas moléculas.

Esto es una aplicación usando machine learning (aprendizaje de máquinas).

Su investigación en la predicción de la energía de enlace para la unión de proteínas se ha enfocado en caracterizar las uniones entre las moléculas conocidas como mhc-i (Major Histocompatibility Complex) y péptidos (secuencias de entre 8 y 11 aminoácidos), los cuales son pequeñas muestras de proteínas celulares o patógenas presentes dentro de la célula. Las moléculas mhc-i se enlazan a los péptidos y los presentan en la superficie de la célula para ser reconocidos por las células T. Si la célula T reconoce el péptido como perteneciente a una proteína patógena, la célula infectada es destruida.

Existen más de 1000 tipos diferentes de moléculas mhc-i y cada tipo en particular se puede enlazar solo a un cierto grupo de péptidos, por lo tanto estimar que tan factible es que un cierto tipo de molécula mhc-i se pueda enlazar a cierto péptido es de vital importancia para el desarrollo de vacunas de ADN.

El modelo para la estimación de la energía de enlace entre moléculas mhc-i y péptidos se realiza usando estructuras tridimensionales de complejos de enlace entre un par determinado de molécula mhc-i y péptido, las cuales se obtienen mediante técnicas de cristalografía. Con esta información se determinan los aminoácidos en las moléculas mhc-i que interactúan con algunos aminoácidos del péptido, esta información, así como ejemplos de pares de moléculas mhc-i-péptidos con su respectiva energía de enlace (medida en el laboratorio) es lo que se necesita para aprender el modelo. Todo lo demás se aprende usando técnicas de machine learning.

Es importante recalcar que existen ejemplos de pares de moléculas mhc-i-péptidos con su respectiva energía de enlace solo para cerca de 50 tipos diferentes de moléculas mhc-i aunque existen más de 1 000 tipos. Grupos de investigación que han atacado este problema en el pasado han construido modelos (usando redes neuronales o support vector machines) que son específicos para el tipo de molécula mhc-i de tal manera que solo pueden predecir energías de enlace para aquellos tipos de moléculas mhc-i para los cuales tiene ejemplos para diseñar el modelo. En el caso del modelo propuesto por el doctor Reyes Gómez, el modelo aprende un grupo de parámetros globales que se puede utilizar para cualquier tipo de molécula mhc-i, de tal manera que aunque el modelo es aprendido utilizando

información de energías de enlace de tan solo 50 tipos de moléculas mhc-i, una vez aprendido el modelo puede estimar energías de enlace para cualquiera de los 1 000 tipos de moléculas mhc-i.

El modelo se usa como herramienta para analizar patrones de mutación en virus, que en el caso del vih las mutaciones son aleatorias y muy rápidas, aunque el modelo no explica directamente el por qué de los cambios. Las mutaciones que permanecen son las que le dan una ventaja evolutiva al virus. Cuando un aminoácido muta y ese hecho debilita al virus, este tipo de mutaciones eventualmente desaparece, según las leyes de la selección natural de Darwin. El modelo puede ser utilizado para explicar el carácter de dichas ventajas o desventajas.

Cada persona tiene seis tipos (de los más de 1 000 posibles) diferentes de moléculas mhc-i, lo cual define de gran manera la unicidad del sistema inmunológico de cada persona. Aquí se manifiesta por qué es importante desarrollar un modelo que pueda estimar energías de enlace para cualquier tipo de molécula mhc-i y no solo de las 50 para las que hay mediciones de energía de enlace disponible, ya que es muy probable que un paciente cualquiera tenga un tipo de molécula mhc-i que no se encuentre dentro del reducido grupo de 50 tipos para los cuales hay mediciones de energía de enlace.

Este nuevo modelo ha sido utilizado para analizar patrones de mutación en las secuencias de aminoácidos en proteínas del vih en pacientes infectados con dicho virus. El análisis se realiza utilizando muestras tomadas con más de 18 meses de diferencia. Gracias a que el modelo puede generalizar a cualquier tipo de molécula mhc-i, el análisis de cada paciente es personalizado ya que se realizan estimaciones de energía de enlace utilizando únicamente los tipos de moléculas mhc-i presentes en cada paciente. Ahora, gracias al trabajo del equipo del doctor Reyes Gómez, se está en la posibilidad de analizar por qué ciertas mutaciones permanecen y de este modo justifican su ventaja sobre otras. Ahora tenemos una herramienta útil para que los biólogos lleven a cabo sus investigaciones de una manera más asertiva y sencilla.

Parte del trabajo que se está haciendo es contribuir al diseño de una vacuna más eficaz contra el vih, tal vez mediante un análisis del genoma de cada individuo para desarrollar la vacuna. En ese sentido, la herramienta desarrollada por el equipo del doctor Manuel Reyes Gómez será indispensable, porque ayudará a predecir el comportamiento individual frente a cierto tipo de vacunas, en caso de que se necesitara de vacunas especializadas. De hecho, se han llevado a cabo pruebas clínicas en hospitales en la Universidad de Harvard donde péptidos provenientes de secuencias de proteínas del vih interactúan con la sangre de pacientes y se comprueba su reacción ante el sistema inmunológico, lo que depende en parte de las moléculas de mhc-i de cada paciente. Se ha podido predecir con cierta certeza si eso ocurrirá o no con este método.

El doctor Reyes Gómez también trabaja en detectar patrones de expresión en ciertos genes que pudieran llevar a tratamientos alternativos para diversos tipos de cáncer.

El doctor Manuel Reyes Gómez, investigador de Microsoft Research, es ejemplo de cómo un investigador mexicano puede destacar en el extranjero. Originario de Puebla, Puebla, fue finalista de las olimpiadas nacionales de matemáticas representando al estado de Yucatán en 1989. Obtuvo la licenciatura en Ingeniería Electrónica y Comunicaciones en la Universidad de las Américas-Puebla. Trabajó como profesor en la Universidad Autónoma del Carmen y como ingeniero de proyectos en la Corporación Mexicana de Investigación de Materiales. Obtuvo la maestría y el doctorado en Ingeniería Eléctrica en la Universidad de Columbia, Estados Unidos, donde se especializó en el reconocimiento de patrones y aprendizaje de computadoras para aplicaciones en audio.

Realizó estancias de investigación como estudiante en las divisiones de investigación de Mitsubishi y Microsoft. Forma parte del grupo pionero de investigadores en Microsoft Research, donde trabaja en proyectos de bioinformática como el anteriormente descrito.

Su trabajo tanto en procesamiento de audio como en bioinformática ha sido presentado y publicado en diversas revistas de investigación y conferencias internacionales en Europa, Asia y América, así como en dos libros de investigación en este campo del conocimiento.

Es coautor de diversas patentes tanto con Microsoft Research como con Mitsubishi Research Labs

El trabajo del equipo de Microsoft Research se extiende al desarrollo de un modelo probabilístico de todo sistema inmunológico.