
**ARTIFICIAL INTELLIGENCE FOR COMBINATORIAL
BIOMARKERS DISCOVERY**

**CLAUDIO DURAN
INGENIERO EN BIOINFORMÁTICA**

RESUMEN

Los sistemas biológicos complejos que emergen a diferente escala pueden ser representados y analizados como redes, por ejemplo: los ecosistemas pueden ser modelados como redes de especies que interactúan; una vía celular como una red de moléculas que interactúan; y las proteínas como una red de aminoácidos. Nodos y arista son el componente básico de una red, que al mismo tiempo es una herramienta fundamental en la física de sistemas complejos. Los nodos representan características (o unidades) que señalan las partes en que el sistema complejo está organizado como red, mientras que las aristas representan la interacción entre unidades. Por lo tanto, debido a su conveniencia en la representación de la física de sistemas complejos, las redes son hoy en día sumamente usadas no solo en la representación y simulación de sistemas biológicos, sino que también en diversas áreas como, entre otras: ciencias sociales, comercio, administración y economía. Con el fin de inferir redes de características discriminativas directamente de un cohorte e muestra de datos o señales, es importante detectar patrones ocultos que revelen segregación de muestras en diferentes sub-grupos. Para obtener la mejor detección de patrones de muestra se adoptaron métodos de aprendizaje automático no supervisados para ejecutar reducción de dimensionalidad como: análisis de componentes principales (PCA) e incrustación curvilínea mínima (MCE). No obstante, una tarea importante y no trivial es encontrar la combinación óptima de estrategias de reconocimiento de patrones entre la normalización de datos técnicas de reducción de dimensionalidad (lineal y no lineal), número de dimensiones comparadas, detección de grupos y/o evaluación por medio de calificaciones cuantitativas. Después de reconocer los patrones de la muestra, el paso final fue detectar la combinación de características (y sus asociaciones respectivas) que son responsables para la segregación de patrones, y representar sus relaciones en una red, en donde los nodos son característicos. Esta red ayudó a encontrar combinatorias de biomarcadores cruciales (combinación de características) que

pueden justificar las diferencias entre un estado normal y patológico, útil también para diseñar estrategias de detección temprana en pacientes. El proyecto está enmarcado entre física de sistemas complejos, aprendizaje automático y ciencia de redes, en el área de la inteligencia artificial y sistemas complejos. Más específico, este trabajo apunta por el descubrimiento de un enfoque automático y explorativo que, dado un set de datos, es capaz de detectar la mejor combinación de estrategias de reconocimiento de patrones para la detección de biomarcadores basada en un enfoque de redes. Obteniendo interesantes resultados incluso de datos complejos.