

FACULTAD DE INFORMÁTICA

SELECCIÓN DE CARACTERÍSTICAS EN ENTORNOS BIG DATA. APLICACIÓN EN GENE SIGNATURES

Camele, Genaro

Hasperué, Waldo (Dir.)

Instituto de Investigación en Informática (III-LIDI).

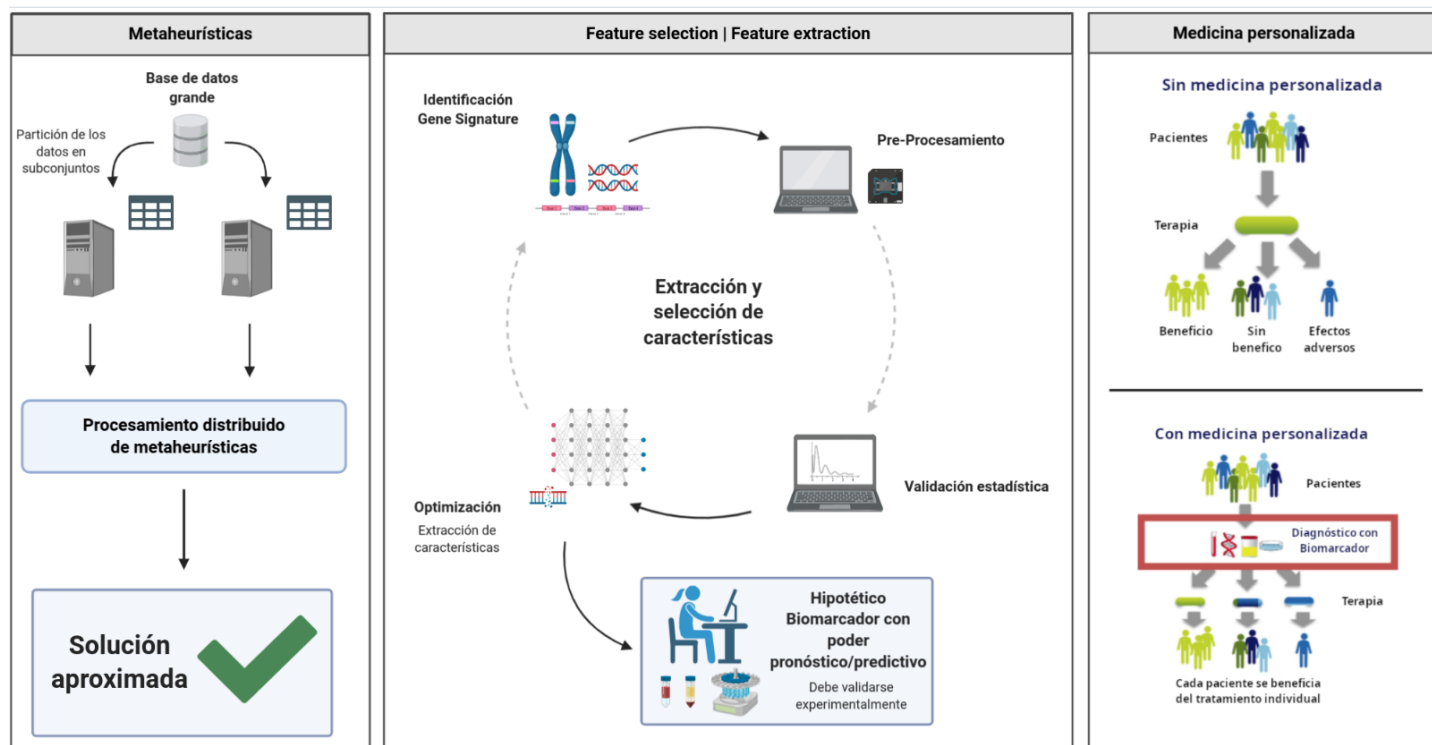
gcamele@lidi.info.unlp.edu.ar

PALABRAS CLAVE: Gene signatures, Biomarcadores, Big Data, Metaheurísticas.

FEATURE SELECTION IN BIG DATA ENVIRONMENTS. APPLICATION IN GENE SIGNATURES

KEYWORDS: Gene signatures, Biomarkers, Biomarkers, Big Data, Metaheuristics.

Resumen gráfico



Resumen

En el área de la minería de datos y su aplicación con técnicas de machine learning, los algoritmos de selección de características juegan un papel muy importante. El objetivo de esos algoritmos es el de reducir las entradas a un tamaño apropiado para su procesamiento y análisis. Cuando el volumen de información a procesar crece hace que la ejecución de los algoritmos de extracción de características convencionales tenga un tiempo de procesamiento sumamente alto. Si bien puede considerarse la separación o el análisis independiente de cada característica, muchas veces resulta útil poder analizar correlaciones entre dos o más variables. La selección de características puede aplicarse a la medicina genómica, la cual ayuda a entender de forma más precisa por qué enfermamos, y el peso que tiene en una enfermedad la existencia de defectos genómicos frente a factores medioambientales que pueden desencadenar una enfermedad concreta. En el ámbito de la genómica funcional, se destaca el análisis de perfiles de expresión génica; éstos tienen como objetivo principal la identificación de un grupo de genes, cuyo patrón de expresión se encuentren asociados a un fenotipo en particular, concepto conocido como gene signature.

Un objetivo particular de los signatures es su utilidad como biomarcador diagnóstico, pronóstico o predictivo de una patología en estudio. Los biomarcadores con valor pronóstico permiten una mejor estratificación de pacientes. Para llevar a cabo el descubrimiento de nuevos gene signatures es necesario un proceso de automatización que permita encontrar genes candidatos en base al conocimiento del experto. En la actualidad esta tarea es realizada de forma manual. Con la rápida acumulación de datos de expresión génica de diversas tecnologías los algoritmos automáticos de reducción de dimensiones pueden seleccionar aquellas que resulten más representativas del conjunto de características. Los resultados de esta selección podrían ser interpretados como un posible gene signature.

El objetivo general de este plan de beca es el de contribuir con el desarrollo de algoritmos de extracción de características en entornos Big Data que permitan la identificación y la evaluación de gene signatures.