INVESTIGACIÓN JOVEN
Revista Electrónica
ISSN: 2314-3991

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). Facultad de Informática, UNLP.

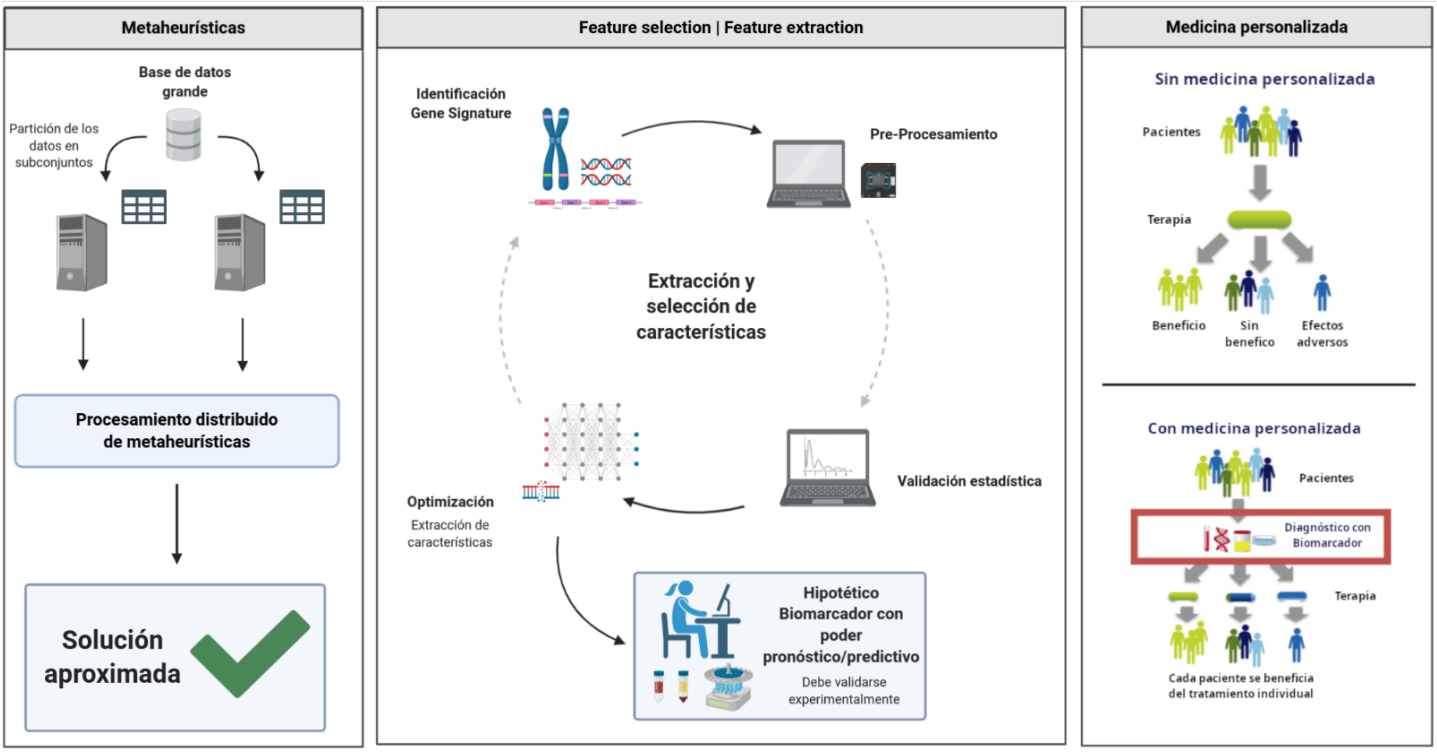
gcamele@lidi.info.unlp.edu.ar

PALABRAS CLAVE: Extracción de Características, Selección de Características, Biomarcadores.

FEATURE SELECTION IN BIG DATA. APPLICATION IN GENE SIGNATURES

KEYWORDS: Big Data, Machine Learning, Feature Selection, Feature Extraction, Gene Signatures.

Resumen gráfico



INVESTIGACIÓN JOVEN
Revista Electrónica
ISSN: 2314-3991

Resumen

La medicina genómica es aquella que utiliza el conocimiento del genoma humano y de ciencias afines para identificar el riesgo de padecer una enfermedad, diagnosticarla precozmente y tratarla de forma personalizada. La medicina genómica 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; estos 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. Estos son un conjunto de genes que se sospecha, podrían ser marcadores de una patología en particular. Para evaluar la eficacia del mismo, se procesa un dataset que consta de pacientes que sufren la patología y la expresión de los genes que están especificados en el gene signature para cada una de estas personas, el tiempo transcurrido desde el último chequeo realizado y el estado vital del paciente.

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 según su pronóstico de progresión de enfermedad independientemente de una terapia, abriendo el paso a investigaciones de tratamientos adecuados para cada categoría de paciente definida. Por otro lado, los biomarcadores con valor predictivo permiten predecir si un tratamiento tendrá o no efecto en un paciente, logrando evitar tratamientos en pacientes para los cuales se supone no tendrán efecto positivo.

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, es posible aplicar algoritmos automáticos de reducción de dimensiones, con el objetivo de seleccionar aquellas que resulten más representativas del conjunto de características. Los resultados de esta selección podría ser interpretada como un posible gene signature.

Numerosos métodos han sido desarrollados para la selección de características en bioinformática. Todavía es un desafío elegir un método apropiado para un problema específico y buscar las características de clasificación más razonable. No obstante, la implementación con algoritmos paralelos no ha sido aún estudiado en profundidad.

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

Multimedia

http://sedici.unlp.edu.ar/handle/10915/114316