

Un rápido diagnóstico de cáncer de mama basado en espectroscopia Raman

A rapid diagnosis of breast cancer based on Raman spectroscopy

Francisco-Javier Luna-Rosas¹, Julio-César Martínez-Romo¹, Ricardo Mendoza-González¹, Huizilopoztli Luna-García², Mario-Alberto Rodríguez-Díaz¹, Laura C. Rodríguez-Martínez¹, Marco-Antonio Hernández Vargas¹

¹ Tecnológico Nacional de México/Instituto Tecnológico de Aguascalientes (México)

² Universidad Autónoma de Zacatecas (México)

DOI: <http://dx.doi.org/10.6036/9047>

En la actualidad más del 99% de los objetos que se encuentran en el mundo físico aún no están conectados a internet, sin embargo, un fenómeno denominado Internet del Todo (IdT), activará todo lo que puedan imaginarse. Para el año 2020 habrá 50,000 millones de objetos inteligentes conectados a internet [1]. IdT modificará la manera de interactuar entre los habitantes de la tierra y con los productos que utilizamos. Las aplicaciones demandarán soluciones innovadoras y eficientes de procesamiento para la mejora del conocimiento y toma de decisiones en las organizaciones, escuelas y hospitales. El volumen de datos que, hace una década, se producía en un año, ahora se produce en una semana. A medida que más objetos no conectados se conectan, los datos siguen creciendo exponencialmente [1]. Como resultado, IdT tendrá cada vez más que ver con la velocidad de procesamiento.

La computación paralela es una forma de cómputo en la que muchas instrucciones se ejecutan

simultáneamente, operando sobre el principio de que problemas grandes, a menudo se pueden dividir en unos más pequeños y luego son resueltos simultáneamente (en paralelo) con el objetivo de hacer más eficiente el procesamiento de las aplicaciones. Con el desarrollo de hardware que soporta la programación paralela, especialmente el desarrollo de computadoras multicore, las arquitecturas de programación paralela llegan a ser más importantes que antes [2]. No es de sorprenderse que el *software* R haya llegado a ser el favorito en la era de Big Data Analytics [2]. Snow provee el soporte para ejecutar fácilmente funciones en R en paralelo utilizando la tecnología multicore. Snow utiliza una arquitectura, maestro/esclavo donde el maestro envía tareas a los trabajadores y los trabajadores ejecutan las tareas y retornan los resultados al maestro. En [3] se presenta una investigación que enlaza los requerimientos futuros de procesamiento de cómputo y las tendencias tecnológicas actuales en redes y computo denominado Internet del Todo (IdT).

En [3] presentamos PCA y SVM en paralelo para optimizar el tiempo de respuesta en la detección automatizada de cáncer de mama, en la cual una señal Raman es clasificada como biopsias de tejido sano (clase ω_1) y tejido dañado (clase ω_2). El objetivo de los métodos de reducción de la dimensión específicamente el Análisis de Componentes Principales (PCA por sus

siglas en inglés) es bajar la dimensión, normalmente en las tablas de datos, en el caso de espectros Raman si utilizáramos el espectro completo tendríamos 766 variables, que corresponden a 766 puntos capturados (R766), lo que haría imposible de visualizar los datos, esta es la razón por la cual aplicamos PCA para reducir las dimensiones, y optimizar la caracterización del espectro. Por otra parte, las Maquinas de Soporte Vectorial (por sus siglas en inglés SVM) son un algoritmo de aprendizaje supervisado. La SVM es entrenada en dos pasos. En el primer paso, una SVM es entrenada como un clasificador con una parte de los datos. En el segundo paso (es decir, predicción), usamos el clasificador entrenado en el primer paso para clasificar el resto de los datos. Después de que la SVM ha demostrado tener buena clasificación, en el artículo se muestra cómo podemos utilizar el procesamiento paralelo para optimizar el tiempo de respuesta de la clasificación de la SVM en espectros de cáncer de mama de tejido sano y dañado. Los resultados demuestran que la combinación de PCA y SVM en paralelo optimizan el tiempo de respuesta en más del 45%, por lo que podemos decir que las implementaciones de estas dos herramientas generan información verídica y útil para proveer a los especialistas herramientas clínicas importantes para una rápida y eficiente detección automatizada de cáncer de mama basada en espectroscopia Raman.

REFERENCIAS

- [1] CISCO netacad. Curso de Cisco en Línea para Internet de las Cosas, Recuperado en Julio 2018. <http://netacad.com/es/courses/intro-iot/>.
- [2] McCallum, Q.E. Weston, S. Parallel R: Data Analysis in the Distributed World, O'Reilly 2012. ISBN 978-1-44930-992-3. Disponible en: <https://itbooks24.com/ebook/parallel-r>.
- [3] Luna-Rosas FJ, Martínez-Romo JC, Mendoza-González R, Luna-García H, Rodríguez-Díaz MA, Rodríguez-Martínez LC, Hernández-Vargas MA. PCA y SVM en Paralelo para Optimizar el Diagnóstico de Cáncer de Mama Basado en Espectroscopia Raman, DYNA New Technologies, Enero-Diciembre 2018, Volumen 5, [14 p.], DOI: <http://dx.doi.org/10.6036/NT8597>

