

El papel de la ciencia de datos y el aprendizaje automático en la Química Analítica

Dr. Eduardo Rodríguez de San Miguel Guerrero

La ciencia de datos es el campo que estudia los datos y cómo extraer significado de ellos, mientras que el aprendizaje automático (ML) es el estudio de algoritmos informáticos que pueden mejorar automáticamente a través de la experiencia y el uso de datos [1]. Los algoritmos de ML se entrenan para hacer clasificaciones o predicciones, discerniendo información clave dentro de los proyectos de minería de datos para impulsar los procesos de toma de decisiones en diferentes campos de aplicaciones.

Actualmente se han desarrollado herramientas analíticas innovadoras para el análisis de muestras diversas como sensores multiplexados, cromatografías acopladas, cromatografías multidimensionales, espectroscopias simultáneas, así como diversas técnicas de imagen, que han demostrado un rendimiento excepcional para proporcionar una mayor sensibilidad y resolución para abordar problemas complejos que antes no se podían resolver con las herramientas analíticas tradicionales. Por otro lado, estas nuevas técnicas generan datos mucho más grandes y complejos que los métodos tradicionales de análisis de datos pueden no ser capaces de interpretar.

Los diferentes aspectos de los grandes datos en Química Analítica pueden también ser abordados por nuevos desarrollos en diferentes dominios de datos científicos, dado que los desafíos y métodos son muy similares y, a menudo, genéricos [2]. La calidad de los datos, los métodos eficientes de visualización de ellos y la validación de modelos son aspectos críticos comunes en el análisis de grandes conjuntos de datos, analíticos o no. Así ante los grandes retos analíticos impuestos por en ciencia forense, la seguridad alimentaria y la caracterización de materiales, por citar algunas áreas, las técnicas convencionales de análisis de datos para la detección de señales y la identificación de compuestos en Química Analítica se encuentra en un proceso de ampliación a través del uso de algoritmos basados de ML, como una mejor alternativa a los enfoques convencionales basados en umbrales, para la detección dirigida y no dirigida de diversas sustancias químicas.

En esta plática se proporciona una descripción general de las metodologías de minería de datos en la Química Analítica actual, los enfoques y los métodos utilizados y se ejemplifica el uso de técnicas de ML en aplicaciones características.

[1] T.M. Machine Learning. New York: McGraw Hill (1997). ISBN 0-07-042807-7.

[2] Szymanska, E. Modern data science for analytical chemical data: A comprehensive review. Anal. Chim. Acta 1028 (2018) 1-10.