



Análisis del Comportamiento Dinámico de los Patrones de Precipitación en la provincia de Los Santos de Panamá

Javier E. Sánchez-Galán (javier.sanchezgalan@utp.ac.pa)^{1,*}, Marlemys Martínez (marlemys.martinez@utp.ac.pa)², Vicente Navarro (vicente.navarro@utp.ac.pa)², Romnan Gordon (gordon.roman@gmail.com)³, Tomas Morales Acotlzi (acotlzi@atmosfera.unam.mx)⁴

¹Facultad de Ingeniería de Sistemas Computacionales, Universidad Tecnológica de Panamá

²Facultad de Ciencias y Tecnología, Universidad Tecnológica de Panamá

³Instituto de Innovación Agropecuaria de Panamá (IDIAP),

⁴Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México

*Autor de correspondencia

Palabras Clave: *Descriptores Moleculares, Productos Naturales, Redes Neuronales Atencionales, Redes Mensajeras, Redes Neuronales Recurrentes.*

Una de las formas de identificar sin ambigüedades las estructuras de compuestos químicos y, particularmente productos naturales, es mediante la notación lineal de SMILES canónicos. Por su naturaleza textual y lineal, guardan información espacial y de algunos descriptores en una cadena ininterrumpida de caracteres, convirtiéndose en una representación estructural muy frecuente en quimioinformática. Sin embargo, esta información también pertenece a un conjunto representable por medio de grafos, a lo cual pueden existir propiedades interconectadas (como toxicidad, actividad, solubilidad o energía de solvatación) que no sean apropiadamente identificadas ni predichas usando solo Redes Recurrentes. Peor aún si hablamos de relacionar propiedades continuas con propiedades multicategóricas.

En el presente trabajo se busca explorar diversas modificaciones a la arquitectura propuesta por la red SMILES2vec, capaces de mejorar su desempeño para valores etiquetados con más de dos categorías. La quimioteca a estudiar NAPROC-13 y se pretende predecir el tipo de compuesto deuterado por el cual puede (en los casos de quimioprospección) o pudo disolverse (en los casos no reportados) un producto natural, al realizarse la resonancia magnética nuclear (NMR) respectiva. Se evalúan cadenas de hasta 160 caracteres, representando el 96,1% de los productos naturales enlistados en la base de datos, curados y en su representación quiral. Luego se determinan los descriptores constitucionales, propiedades de interés y huellas moleculares para caracterización del espacio químico mediante la librería RDKit en Python y el software Molecular Operating Environment, a fin de seleccionar aquellos descriptores más determinantes para la diferenciación de solventes a predecir y los cuales se implementarán para las mejoras, contemplándose arquitecturas propias de Redes con Mecanismos de Atención (GAT) y Mensajeras (MPNN).



MEMORIAS DEL XXXI CONGRESO MEXICANO DE
METEOROLOGÍA Y XVI CONGRESO INTERNACIONAL DE METEOROLOGÍA,
Año 2023, No. 22. ISSN: 2594-1836

Entre las técnicas para evaluar su desempeño se emplean la Media Cuadrática para los valores continuos, y Entropía Cruzada Categórica y Precisión Categórica para los valores en categorías, con lo que previamente los datos en formato SMILES se codificarán como vectores one hot. Los resultados, en comparación a las Redes Recurrentes, muestran mejoras de precisión en la predicción de hasta diez puntos porcentuales.