

Propuesta de Proyecto de Investigación
Maestría en Ciencias y Tecnologías de la Información

22 de octubre del 2021

1. Nombre del proyecto: Generación de estructuras moleculares mediante aprendizaje por refuerzo profundo.

2. Responsable(s)

Dr. Jorge Juarez Gómez

Universidad Autónoma Metropolitana, laboratorio R-105, jig85@xanum.uam.mx

Dr. Eric Alfredo Rincón García

Universidad Autónoma Metropolitana, cubículo T-143, rincon@xanum.uam.mx

3. Área(s) de conocimiento relacionada(s) con el proyecto

Optimización e inteligencia artificial

Química Analítica

Observación: Es recomendable que el estudiante tenga conocimientos de redes neuronales.

4. Descripción del proyecto

- Contexto

Aprendizaje por refuerzo profundo (ARL) es una herramienta de inteligencia artificial que combina técnicas de aprendizaje por refuerzo con redes neuronales, que busca entender las reglas que rigen el problema de interés mediante un proceso de prueba y error. Algunos de sus logros más llamativos se encuentran relacionados con aplicaciones en juegos de mesa como go, ajedrez, backgammon, entre otros, y juegos de consola, donde han sido capaces de superar el desempeño humano. Sin embargo, ARL ha demostrado gran versatilidad y se han explorado aplicaciones en el manejo automatizado de vehículos, predicciones en economía y finanzas, sistemas de recomendación, por mencionar algunos casos. En particular, para este proyecto es de especial interés la generación de nuevas moléculas, como se ha reportado en [1, 2, 3].

- Motivación

El aprendizaje por refuerzo profundo requiere de sólidos conocimientos en programación y en redes generativas adversarias. Por lo tanto, se convierte en un área de interés para cualquier persona interesada en aprendizaje automatizado. Además de que su uso para la generación de nuevas moléculas ha despertado un interés en diversa áreas relacionadas con Química.

- Aporte esperado al área de conocimiento

Se analizará el desempeño de un algoritmo basado en redes neuronales para generar moléculas, por lo que los resultados pueden ser muy valiosos para dos áreas de investigación, inteligencia artificial y química analítica. Además, se espera que las moléculas propuestas puedan ser utilizadas en el diseño de nuevos sensores químicos para la cuantificación de contaminantes.

5. Objetivos

Objetivo general

Generar estructuras moleculares mediante un algoritmo basado en aprendizaje por refuerzo profundo.

Objetivos particulares

- Estudiar los conceptos más importantes sobre aprendizaje por refuerzo y redes neuronales básicas.
- Estudiar los conceptos más importantes sobre redes neuronales generativas adversarias.
- Revisar el estado del arte para seleccionar el tipo de algoritmo más adecuado al problema.
- Seleccionar las características de las moléculas que se generarán.
- Realizar el entrenamiento del algoritmo.
- Validar teóricamente las estructuras moleculares generadas.
- Reportar los resultados obtenidos en la Idónea Comunicación de Resultados (ICR).

6. Metodología

- Estudiar los conceptos más importantes sobre aprendizaje por refuerzo, redes neuronales básicas, y redes neuronales generativas adversarias.
- Revisar el estado del arte para seleccionar las redes neuronales adecuadas al problema.
- Desarrollar y entrenar el algoritmo basado en las redes seleccionadas para la generación de las moléculas con las características deseadas.
- Analizar el desempeño de los algoritmos propuestos mediante la validación teórica de las moléculas generadas.
- Reportar los resultados en la ICR.

7. Calendarización de actividades

Trimestre 1: Estudio de aprendizaje por refuerzo profundo y redes neuronales.

Trimestre 2: Desarrollo y entrenamiento del algoritmo para la generación de moléculas.

Trimestre 3: Validación teórica de las moléculas generadas y ajustes al algoritmo. Entrega de la versión final de la Idónea Comunicación de Resultados.

Trimestre 4: Revisión de los sinodales de la Idónea Comunicación de Resultados. Presentación del examen de grado.

8. Infraestructura necesaria y disponible

Una computadora con Windows o Linux con tarjeta de video Nvidia y Python para programar.

9. Lugar de realización

El proyecto puede realizarse en el cubículo T-145, o de forma remota (no presencial) dependiendo de las circunstancias.

10. Entregables

Idónea Comunicación de Resultados.

11. Referencias bibliográficas básicas

- 1) Born, J., Manica, M., Oskooei, A., Cadow, J., Markert, G., Rodríguez Martínez, M., (2021). PaccMann^{RL}: De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. IScience, Vol. 24 (4). Disponible en: <https://www.sciencedirect.com/science/article/pii/S2589004221002376>
- 2) Goel, M., Raghunathan, S., Laghuvarapu, S., Priyakumar, U. D., (2021). MoleGuLAR: Molecule Generation using Reinforcement Learning with Alternating Rewards. Theoretical and Computational Chemistry. Disponible en: <https://chemrxiv.org/engage/chemrxiv/article-details/60fd7834d03b3d13a602dbfc>

- 3) Pereira, T., Abbasi, M., Ribeiro, B., Arrais, J. P., (2021). Diversity oriented Deep Reinforcement Learning for targeted molecule generation. *Cheminform*, Vol. 13. Disponible en: <https://jcheminf.biomedcentral.com/articles/10.1186/s13321-021-00498-z>