

Cómo diseñar la Red Neuronal más poderosa

Andrea Manuel Simón y Ana López Palomo

November 2023

1 Introduction

Graph Isomorphism Network (GIN) es un modelo de aprendizaje profundo diseñado para clasificar gráficos. Se creó para hacer tareas topológicas de grafos. Además es muy eficaz en la clasificación de moléculas, documentos y otros datos estructurados en forma de grafos.

GIN destaca por su habilidad para reconocer patrones estructurales que no varían en la permutación de nodos. Utiliza capas de agregación que permite la iteración sobre la información de los vecinos de un nodo para actualizar la representación del nodo en cada capa, con el fin de capturar información global del grafo.

Contiene dos operaciones: la operación de agregación, es sensible al orden de los nodos, haciendo que mejore la capacidad del modelo para capturar la información de la estructura del grafo; y la operación de reducción de cardinalidad, mejora la invarianza del modelo en la permutación de nodos. Es esencial en la clasificación de tareas donde el orden de los nodos no es importante.

1.1 Relevancia de GIN

GIN es relevante en:

- Análisis de redes sociales, se utiliza para modelar la estructura de la red y predecir la categoría de nodos, como la clasificación de perfiles de usuarios.
- Aplicaciones en química computacional, donde la estructura molecular es crítica para la actividad biológica y sus propiedades.
- Procesamiento del lenguaje natural (PLN), se utiliza para solucionar algunos problemas donde los datos se pueden representar en grafos, como en la clasificación de documentos basada en la estructura de las relaciones entre palabras y oraciones.
- Biología computacional: se usa para clasificar estructuras moleculares, redes de interacciones genéricas o para predecir funciones de proteínas basándose en la información topológica del grafo.

2 Implementación

En este apartado vamos a explicar el código de python que está en el repositorio, y daremos conclusiones.

Tras cargar y explorar los datos, hemos creado una gráfica aleatoria que representa las proteínas. Luego, hemos preparado los conjuntos de entrenamiento, validación y prueba. Vemos que las proteínas no son un gran conjunto de datos, podríamos usar perfectamente una GNC o un GAT, pero utilizaremos una red de isomorfismo gráfico (GIN). Además, la compararemos con un GCN.

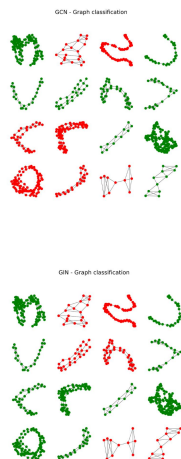
En los resultados del entrenamiento, podemos ver que no hay comparación entre un modelo y otro, con una brecha de un 10 sobre 100 en el promedio de la precisión. Esto se debe a varias razones:

1. El agregador GIN está diseñado específicamente para discriminar gráficos que el agregador GCN no puede.
2. Los vectores ocultos del gráfico de cada capa se concatenan en lugar de simplemente considerar el último.
3. El operador suma es superior al operador medio, al menos en teoría.

Por último, hemos agrupado en un mismo conjunto las incrustaciones de gráficos, tomando la media de los vectores de salida normalizados. Al juntarlos, hemos mejorado la precisión ligeramente.

3 Visualización y análisis de resultados

Del código y las gráficas que hacemos en el código, llegamos a las siguientes conclusiones:



Podemos ver que los dos modelos se equivocan en errores diferentes (donde el color verde es el correcto y el rojo no), pero esto es normal en el aprendizaje

automático al utilizar dos algoritmos distintos en el mismo conjunto de datos. La primera imagen representa la solución del modelo GCN, mientras que la segunda es el GIN. Hay una clara diferencia en el error que cometen, ya que la segunda tiene muchos más grafos correctos que la primera.

4 Conclusión

En resumen, la implementación y comparación de GIN con GCN revela aspectos interesantes sobre la eficacia de los modelos. GIN demuestra ser una muy buena opción para la clasificación de grafos, sobre todo en casos donde la invarianza a la permutación de nodos es crucial.

Por otra parte, el diseño del agregador de GIN permite discriminar gráficos de manera más efectiva que el agregador en GCN, y el operador suma en GIN se alinea mejor con los requisitos teóricos de la tarea de clasificación.

Por último, aunque GIN destaca teóricamente y en resultados experimentales, su implementación efectiva en el mundo real puede depender de muchos factores, haciéndola más complicada de aplicar.