

Autoatención para las GNN

Jose Luis Rodríguez, Gonzalo Martínez y Alexandre Muñoz

December 2023

1 Introducción

Las Redes Neuronales Gráficas, o *Graph Neural Networks* (GNN) en inglés, es un enfoque del *Deep Learning* que permite ser aplicado directamente sobre gráficos. Las GNN aprovechan la dependencia de los nodos del grafo con sus vecinos en base a las aristas que los relacionan, consiguiendo inferir así un vector de características computando la información obtenida de los nodos vinculados a él.

Las Redes Neuronales Gráficas emplean las matrices de propiedades y de adyacencia para identificar los nodos vecinos de aquel elemento del que se desea obtener el vector de características aprendido, en el cual influirán las características y tipo de unión de los nodos conectados a él de forma directa. En base a las características inferidas en cada nodo y a la identificación de patrones que relacionan los vínculos entre los elementos del grafo, las GNN podrán aplicarse con distintas finalidades:

- **Clasificación del grafo:** la finalidad de dicha aplicación es clasificar el conjunto completo del grafo entre distintas categorías, siendo una de las aplicaciones más empleadas en el sector químico y biomédico con el objetivo de clasificar estructuras moleculares.
- **Clasificación de nodos:** el objetivo será determinar el vector de propiedades inferido en cada nodo en base a la información obtenida de las entidades vecinas. Se entrena empleando parte del grafo sin etiquetar, consiguiendo así que nuestro modelo aprenda las características dependientes de los nodos vecinos que determinarán la pertenencia del nodo a una cierta categoría. En el sector comercial se emplea por ejemplo con el objetivo de determinar si un usuario de nuestra red comprará o no un cierto producto, siendo etiquetado como potencial comprador o no.
- **Predicción de vínculos:** el algoritmo buscará relaciones entre los nodos vinculados entre sí con el objetivo de predecir nuevas aristas en el grafo. Su aplicación más extendida tiene lugar en el campo de las redes sociales en el cual en base a las características y amistades de un usuario, se realizarán recomendaciones de otros perfiles.

1.1 Capa de atención en gráficos

Las redes de atención gráfica (GAT) son uno de los tipos más populares de redes neuronales gráficas.

En lugar de calcular pesos estáticos basados en grados de nodos como Graph Convolutional Networks (GCN), asignan pesos dinámicos a las características de los nodos a través de un proceso llamado autoatención. La idea principal detrás de los GAT es que algunos vecinos son más importantes que otros, independientemente de sus grados de nodo.

Son arquitecturas de redes neuronales novedosas que operan con datos estructurados en gráficos, aprovechando capas de atención propia enmascaradas para abordar las deficiencias de métodos anteriores basados en convoluciones de gráficos o sus aproximaciones. Al apilar capas en las que los nodos pueden atender las características de sus vecindarios, permitimos (implícitamente) especificar diferentes pesos para diferentes nodos en un vecindario, sin requerir ningún tipo de operación matricial costosa (como inversión) o depender de conocer el gráfico. estructura por adelantado. De esta manera, abordamos simultáneamente varios desafíos clave de las redes neuronales de gráficos basadas en espectros y hacemos que nuestro modelo sea fácilmente aplicable a problemas inductivos y transductivos.

De forma general, los vecinos de un nodo son tratados igualitariamente durante la fase de agregación, en la cual puede ser aplicada la función de media, tratándose de igual forma a todos los nodos a la hora de realizar la suma y la división entre el número de elementos. Sin embargo, al tratar con grafos cuyos nodos representan entidades con distintas propiedades dentro del conjunto, es común que no todos los nodos sean igual de influyentes por lo que algunos de ellos deberían ser ponderados en mayor medida, prestando así más “atención” a los nodos identificados como influyentes. Cada arista tendrá asociada un coeficiente de atención α_{ij} indicando la importancia del nodo ‘j’ para el vector de características inferido del nodo ‘i’. Los coeficientes α asociados a las aristas unidas directamente a un nodo sumarán uno. El mensaje agregado dependerá de los distintos coeficientes haciendo que los nodos cuya influencia se determine como baja o nula tendrán una importancia proporcional a dicha categoría reduciendo su contribución. En el caso de que un nodo tenga contribución de dos vecinos, a pesar de que se aplique la operación de media entre los vectores de propiedades, el valor del vector agregado tenderá a estar mucho más próximo al vector asociado al nodo cuya arista indica que es necesaria una mayor atención.

Para poder aplicar el mecanismo de atención a las predicciones del grafo completo, será necesario determinar la aportación de cada nodo sobre sus vecinos, es decir, será necesario determinar los coeficientes de atención de cada arista. Se concatenarán los vectores de propiedades transformados del nodo actual y del nodo vecino, obtenidos tras multiplicar la matriz de pesos $W^{(k)}$ por el vector de propiedades de la capa actual $\mathbf{h}_i^{(k)}$. Una vez concatenados los vectores transformados, se emplearán como entrada de una neurona simple cuyo peso se denomina $W_a^{(k)}$ y la función de activación asociada será una función LeakyReLU. Finalmente, se aplicará la función Softmax para normalizar los

coeficientes de atención, obteniendo así un coeficiente por cada arista existente entre cada pareja de nodos. La matriz de pesos $W_a^{(k)}$ se ajustará durante el entrenamiento, permitiendo obtener los coeficientes de atención apropiados.

Estos coeficientes se integrarán en la matriz de adyacencia A , donde los valores normalizados serán sustituidos por los coeficientes de atención calculados para cada arista indicada en la matriz.

1.2 Contraste de GAT con otras arquitecturas como GCN

Las redes de atención gráfica (GAT) y las redes de propagación de grafos (GCN) son dos arquitecturas fundamentales en el campo de las redes neuronales convolucionales para grafos (GNN). A continuación, se presenta un contraste más detallado entre GAT y GCN:

GAT (Redes de Atención Gráfica):

- Las GAT utilizan mecanismos de atención para asignar pesos a los nodos vecinos en función de su relevancia para el nodo central en un grafo. Este mecanismo de atención permite que cada nodo del grafo tenga una ponderación diferente para sus vecinos, lo que les otorga mayor flexibilidad en la representación de relaciones entre nodos.
- Al utilizar la atención, las GAT pueden capturar relaciones no lineales y modelar de manera más precisa las interacciones entre nodos en un grafo.
- Las GAT han demostrado ser efectivas en tareas de aprendizaje en grafos, como clasificación y predicción en grafos, y han mostrado un rendimiento competitivo en comparación con otras arquitecturas de GNN.

GCN (Redes de Propagación de Grafos):

- Las GCN utilizan una operación de propagación de información entre nodos vecinos, lo que les permite capturar la estructura local del grafo. A través de esta operación, las GCN pueden capturar la información de nodos vecinos y actualizar la representación de los nodos en función de su entorno local.
- Aunque las GCN son efectivas en la captura de información local y han sido ampliamente utilizadas en una variedad de aplicaciones, pueden tener limitaciones en la representación de relaciones no lineales y en la asignación de pesos diferenciados a los vecinos de un nodo.

Contraste:

- La principal diferencia radica en el mecanismo subyacente para capturar la información del grafo. Mientras que las GAT utilizan mecanismos de atención para asignar pesos diferenciados, las GCN se basan en la propagación de información entre nodos vecinos.

- Las GAT ofrecen una mayor flexibilidad en la representación de relaciones no lineales y en la asignación de importancia a los nodos vecinos, lo que puede ser beneficioso en tareas que requieren un modelado más preciso de las relaciones en el grafo. Esta flexibilidad en la asignación de pesos diferenciados puede ser especialmente útil en grafos densos o en presencia de relaciones no uniformes.
- Por otro lado, las GCN son eficientes en la captura de la estructura local del grafo y en la propagación de información a través de la red, pero pueden tener limitaciones en la representación de relaciones más complejas y no lineales, lo que podría afectar su capacidad para modelar la complejidad de ciertos grafos.

En resumen, tanto las GAT como las GCN son arquitecturas importantes en el campo de las redes neuronales convolucionales para grafos, cada una con sus propias fortalezas y limitaciones en la representación y modelado de grafos. La elección entre GAT y GCN dependerá de la naturaleza de los datos y las tareas específicas que se estén abordando.

2 Conclusión y Generalidad

2.1 Reflexión sobre la relevancia y aplicabilidad de GAT en diversos contextos

La relevancia y aplicabilidad de las redes de atención gráfica (GAT) en diversos contextos se manifiesta a través de su capacidad para modelar relaciones no lineales y capturar la importancia de los nodos vecinos en un grafo. Esto puede ser especialmente relevante en situaciones donde las interacciones entre elementos son complejas y no uniformes, como en el análisis de datos de redes sociales, sistemas de recomendación personalizados o en la comprensión de estructuras de información no lineales. La flexibilidad de las GAT para asignar pesos diferenciados a los nodos vecinos permite su adaptación a una amplia gama de contextos, lo que las hace aplicables en diversas áreas de estudio y análisis, desde la educación hasta la biotecnología y más allá.

Las GAT han demostrado ser efectivas en tareas de aprendizaje en grafos, como clasificación y predicción en grafos, y han mostrado un rendimiento competitivo en comparación con otras arquitecturas de GNN. Su capacidad para capturar relaciones no lineales y modelar de manera precisa las interacciones entre nodos en un grafo las hace relevantes en una variedad de situaciones del mundo real donde la comprensión de las relaciones complejas es fundamental. Además, su flexibilidad en la asignación de pesos diferenciados a los nodos vecinos las hace especialmente útiles en grafos densos o en presencia de relaciones no uniformes, lo que amplía su aplicabilidad en una variedad de dominios.

En resumen, la relevancia y aplicabilidad de las redes de atención gráfica (GAT) en diversos contextos radica en su capacidad para modelar relaciones no

lineales, capturar la importancia de los nodos vecinos en un grafo y su flexibilidad para adaptarse a una amplia gama de situaciones del mundo real. Su efectividad en tareas de aprendizaje en grafos y su rendimiento competitivo en comparación con otras arquitecturas de GNN respaldan su utilidad en una variedad de aplicaciones, lo que las hace relevantes y aplicables en campos que van desde la ciencia de datos hasta la ingeniería, la medicina y más allá.

2.2 Conclusiones sobre el potencial y las limitaciones de la arquitectura GAT.

La arquitectura de redes de atención gráfica (GAT) presenta un gran potencial en la modelización de relaciones no lineales en datos de tipo grafo, lo que la hace relevante en una amplia gama de aplicaciones, desde el análisis de redes sociales hasta sistemas de recomendación personalizados. Su capacidad para capturar la importancia de los nodos vecinos en un grafo y asignar pesos diferenciados a estas interacciones la hace especialmente útil en situaciones donde las relaciones son complejas y no uniformes.

Las GAT han demostrado ser efectivas en tareas de aprendizaje en grafos, como clasificación y predicción en grafos, y han mostrado un rendimiento competitivo en comparación con otras arquitecturas de GNN. Su capacidad para capturar relaciones no lineales y modelar de manera precisa las interacciones entre nodos en un grafo las hace relevantes en una variedad de situaciones del mundo real donde la comprensión de las relaciones complejas es fundamental. Además, su flexibilidad en la asignación de pesos diferenciados a los nodos vecinos las hace especialmente útiles en grafos densos o en presencia de relaciones no uniformes, lo que amplía su aplicabilidad en una variedad de dominios.

Sin embargo, es importante tener en cuenta que las GAT pueden presentar limitaciones en términos de escalabilidad y eficiencia computacional en grafos muy grandes, lo que puede requerir estrategias adicionales para su implementación efectiva en tales contextos. Estas limitaciones pueden surgir debido a la complejidad computacional asociada con el cálculo de los pesos de atención para cada par de nodos en un grafo grande, lo que puede impactar el rendimiento y la eficiencia en aplicaciones a gran escala.

En resumen, la arquitectura GAT presenta un potencial significativo en la modelización de relaciones no lineales en datos de tipo grafo, lo que la hace relevante y aplicable en una amplia gama de escenarios del mundo real. No obstante, es crucial considerar sus posibles limitaciones, especialmente en términos de escalabilidad y eficiencia computacional, al aplicarlas en entornos con grafos muy grandes o en aplicaciones que requieran un procesamiento intensivo.