

Autoatención para las GNN

Andrea Manuel Simón y Ana López Palomo

November 2023

1 Teoría y Fundamentos

Las redes de atención gráfica (GAT) son uno de los tipos más populares de las redes neuronales gráficas. Están diseñadas para hacer tareas en grafos, como, su aprendizaje y clasificación de nodos en redes complejas. En lugar de calcular pesos estáticos basados en los grados de los nodos como GCN, se asignan pesos dinámicoa llamado autoatención. Con esto conseguimos que la red aprenda y se adapte la importancia de las conexiones entre nodos durante el entrenamiento. La idea principal es que algunos vecinos son más importantes que otros independientemente de los grados de sus nodos, en otras palabras, las GAT permite que el modelo aprenda automáticamente cuáles son los nodos más relevantes para cada tarea. También, estas se implementan con múltiples capas de atención, donde cada capa calcula una matriz de atención diferente permitiendo que el modelo considere distintas perspectivas y niveles de detalle.

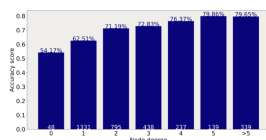
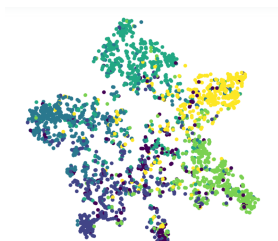
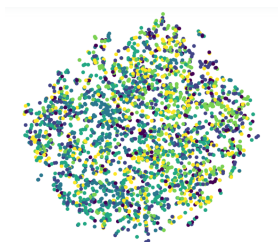
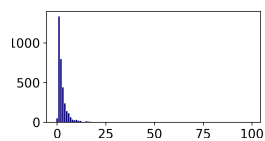
Diferencias entre GAT y GCN:

GCN: los pesos de las conexiones entre nodos son estáticos y se basan en el grado de los nodos. Haciendo que la información de cada vecino contribuye de manera uniforme a la actualización de un nodo. Realiza una operación que suma o promedia las características de los vecinos para actualizar el nodo objetivo. En los grafos grandes puede tener problemas de escalabilidad, al tener la información que dependa de su estructura. La propagación de información se propaga de los nodos vecinos a los nodos objetivos, de forma análoga al filtrado del grafo.

GAT: los pesos de las conexiones entre nodos son dinámicos mediante el mecanismo de autoatención. Realiza una atención selectiva, es decir, algunos vecinos pueden tener una contribución más significativa que otros dependiendo de la tarea. A menudo se implementa con varias capas de atención, permitiendo que podamos capturar información de varios niveles de detalle y perspectiva. Es más escalable en grafos grandes, debido a que cada nodo puede atender selectivamente a sus vecinos, reduciendo la dependencia de la estructura del grafo. Esta atención selectiva nos proporciona una interpretación más fina de las relaciones del grafo, siendo beneficioso en tareas donde hay una gran importancia en las variación de las conexiones.

Además el GAT ofrece una mayor flexibilidad, interpretación y explicabilidad. El rendimiento de ambos depende de la tarea y las propiedades del grafo.

2 Implementación y visualización



La implementación del código se encuentra en el repositorio asociado a este trabajo, igual que las gráficas.

En las imágenes podemos ver que, en el primer grafo, todos los datos están mezclados. Esto es porque la visualización es aleatoria, ya que son productos de matrices de peso.

Tras el entrenamiento, vemos que los nodos que pertenecen a las mismas clases se agrupan. De los colores, sacamos que hay seis grupos, correspondientes a las seis clases de documentos. También se pueden apreciar valores atípicos, que son los colores que no están en su grupo. Aún así, se ve que el GAT es un método bastante eficaz, y relativamente rápido (aunque no tan rápido como los GCN).

3 Conclusión

En resumen, Los GAT, aunque más lentos, son más precisos entrenando que los GCN. El mecanismo de autoatención calcula los pesos de manera automática para producir mejores incrustaciones. Su único problema es que puede ser tedioso si se usan datos masivos, debido a su lento tiempo de entrenamiento.