TP3 Deep Learning in Practice

Zakariae EL ASRI
MSC AI at CentraleSupelec
zakariae.elasri@student-cs.fr

Nicolas GREVET
MSC AI at CentraleSupelec
nicolas.grevet@student-cs.fr

1. Introduction

In this exercise, we implemented a GAT (Graph Attention Network) that is a neural network architecture operating on graph-structured data, leveraging masked self-attentional layers.
GAT is built on the concepts of Self Attention and Multi-Head Attention introduced in the GAT paper while using a different attention mechanism.

2. Implementation:

We constructed a model by stacking 3 GAT layers presented in the GAT paper (2 hidden layers + 1 output layer). The hidden layers consist of K = 8 attention heads computing F= 256 features, followed by an ELU nonlinearity. The output layer consists of K=4 attention heads computing 121 features, followed by a logistic sigmoid activation.
We trained our model on 150 epochs to minimize cross-entropy on the training nodes using the Adam SGD optimizer and a batch size of 2 graphs.
<u>Our model is described in the diagram below:</u>
We perform on each input node a first GAT layer that will transform our input features from 50 to 256*8, then each hidden layer takes as input dimension the output of the previous layer and as output 256*Multiheads. The output layer will have 256*8 as input and 121*4 as output.
We note that in the original paper, they concatenate in hidden layers and average in the last output layer (in this case the output dimension would be 121). However, In the GAT Layer of DGL library, we do a projection then addition in all layers. Hence the output dimension 121*4

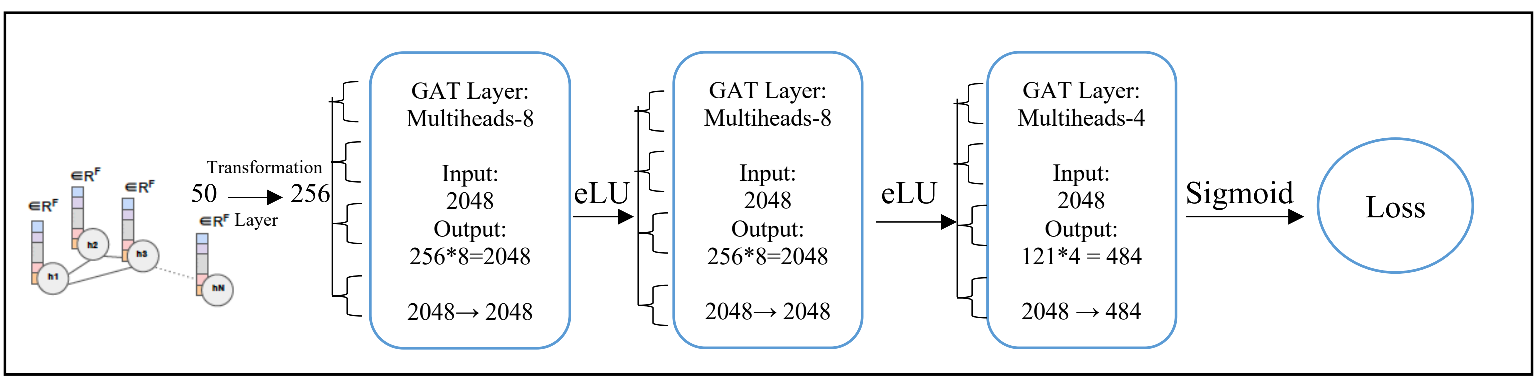

Figure1: Diagram of GAT model with shape information

3. Results:

We tested our model on the nodes of the two test graphs, we had a micro-averaged F1 score of 0.9873

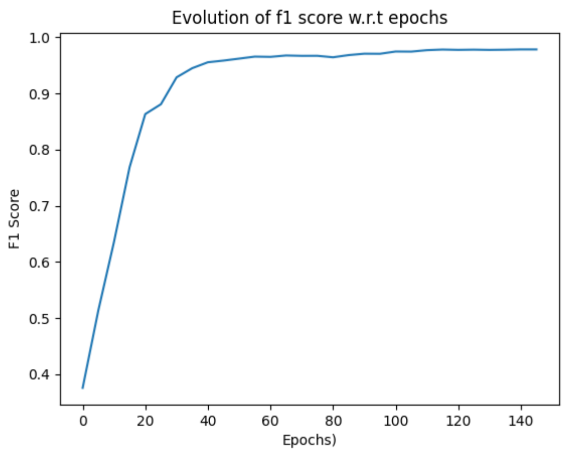

Figure 2: F1-Score evolution w.r.t epochs (Our model)

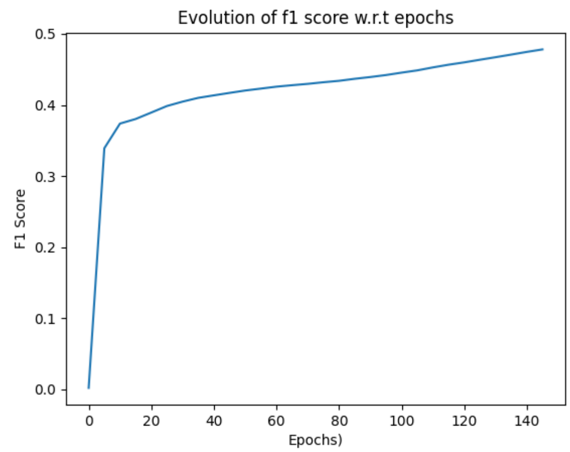

Figure 3: F1-Score evolution w.r.t epochs (Baseline model)

4. Discussion:

F1-score is a metric that combine mixture of precision and recall by calculating their harmonic mean. Our F1-score is equal to 0.9873 (near to 1). This means that both the precision and recall of our model are near to 1. So, our model performs well on classifying each class (It is enough confident) without confusion with other classes.

The GAT model outperforms the ConvGraph for several reasons:

- The GAT allows for assigning different importances to nodes of a same neighborhood, enabling a leap in model capacity, as opposed to GCNs that consider all the neighbors in the same way.
- The attention mechanism in GAT is applied in a shared manner to all edges in the graph, and therefore it does not depend on upfront access to the global graph structure. Also, thanks to this mechanism, the graph is not required to be undirected