Knowledge Lab

Birkbeck UNIVERSITY OF LONDON

The application of Graph Neural Networks and Knowledge Graphs to Education

Research Student: Kristof Neys
Supervisors: George Magoulas, Alex Poulovassilis

1. Research Aim

Our research aim is an investigation of how deep learning methods, such as *Graph Neural Networks* (GNN), *Knowledge Graphs* and *Language models* (Transformers) can be used to analyse a labeled data set obtained from a *Community of Inquiry* experiment performed by Papanikolou et al. [1]. This will allow us to automate the analysis of a student's critical thinking and quality of interaction in online discussions for empowering communities of inquiry.

2. Research Methodology

We deploy three major Deep Learning methods to first do text analysis, followed by a mapping of topics in a Knowledge Graph and finally learning from such a graph using a Graph Neural Network.

2.1 Transformers

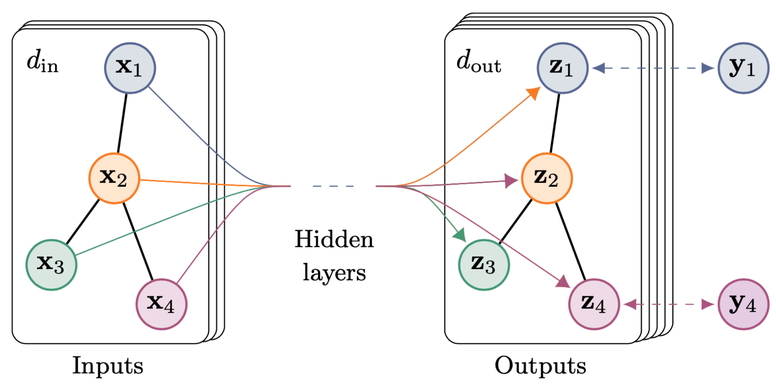
The Transformer model, which is a language model, was developed by Vaswani et al, in 2017 in their seminal paper; "Attention is all you need" [2]. In essence, a transformer model is a deep learning model that can handle sequential input data, such as natural language. However, unlike recurrent neural networks, transformers do not necessarily process the data in order. Instead it deploys an *attention mechanism*, which provides context for any position in the input sequence. For example, in a natural language sentence, the transformer does not need to process the beginning of the sentence before the end. Rather, it identifies the context that provides meaning to each word in the sentence.

2.2 Knowledge Graphs

For our purposes we define a knowledge graph to be:
A structured representation of facts, consisting of entities, relationships and semantic descriptions.
Within this context, entities are real-world objects where the interdependence between entities is reflected in their relationship. Both the entities and relationships are characterised by type and properties. In this research project we will apply the area of Knowledge Graphs to the area of pedagogical content creation. As such, the 'entities' in our definition can be study objects, such as mathematical definitions, historical figures, etc. That is, the entities are the concepts in any learning material. Knowledge graphs are by their nature incomplete and the application of Graph Neural Networks to Knowledge Graphs allows for the discovery of missing links, termed *Knowledge Graph Completion*, thereby establishing new facts.

2.3 Graph Neural Networks

A key component of Graph Neural Network is the concept of *Message Passing Network* (MPN). The essential operation of a MPN is to learn a **richer** representation of a node, that is an *embedding* by incorporating its own feature information, as well as the information from its neighbours, which can be illustrated as follows:



- As shown, there are 4 nodes and the $\mathbf{x}_i$ are the input features (in vector format) of node i. As such our inputs to the MPN are: 1) feature vectors, 2) neighbours represented by an adjacency matrix A.
- **Training:** In a MPN the task is to **aggregate** information from the nodes' neighbours. These are the feature vectors of the neighbours. This operation is performed for every node in the graph.
- Any **forward pass** computes $\mathbf{h}_i^{(k+1)}$ by taking in $\mathbf{h}_i^{(k)}$ and the aggregrated feature information of the neighbours.
- **Learning node embeddings.** Achieved by determining a loss function: a cross-entropy loss on all the **labeled** nodes. Such that:

$$Z = f(Z, A) = \text{softmax}(\hat{A}\ \text{ReLU}\left(\hat{A}XW^{(0)}\right)W^{(1)})$$

where W are learned weight matrices.
- Finally, we optimize the MPN for the task of semi-supervised learning, for instance, node classification using the cross-entropy loss on all *labeled* nodes:

$$\mathcal{L}_{sup} = -\sum_{l \in \mathcal{Y}_L} \sum_{f=1}^{d_{out}} Y_{l,f} \ln Z_{lf}$$

- The neural network weights, $W^{(0)}, W^{(1)}$ are trained using gradient descent. Batch gradient descent is used for the full dataset for every training iteration.

3. Research Approach

We are building a transformer model which analyses the comments that student's wrote during the experiment and that were manually labeled in the study [1]. The transformer model will produce vectors of the written comments which we represent as nodes with feature vectors in the graph together with it's ground-truth label as obtained from the manual classification of the experiment. Having now constructed such a Knowledge Graph, we can train a Graph Neural Network to learn from the various interactions between students and the posted comments. This trained graph model can now be deployed to new experiments, with unlabeled data. Papanikolaou et al. have shown that social network metrics have a positive relationship on Communities of Inquiry. The use of a Graph Neural Network model will allow to discover more sophisticated metrics and *"their impact on the development and evolution of a learning community*[1].

References

[1] Kyparisia Papanikolaou, Maria Tzelepi, Maria Moundridou, and Ioannis Petroulis. Employing social network analysis to enhance community learning. In *International Conference on Intelligent Tutoring Systems*, pages 342–352. Springer, 2020.

[2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.