Heterogeneous Attention Based Modelling for Knowledge Graph Embeddings

Ayushi Agarwal
Department of Computer Science
UCLA
UID: 705496407
ayushi15@ucla.edu

Harini Suresh
Department of Computer Science
UCLA
UID: 505712718
sharini16@g.ucla.edu

Shardul Shailendra Parab
Department of Computer Science
UCLA
UID: 205525006
shardulparab@g.ucla.edu

Abstract

CompGCN Vashishth et al. [2019] is a novel Graph Convolutional Framework which uses composition operations from Knowledge graph Embeddings and also scales with the number of relations. It generalises several of existing multi-relational GCN methods as well. Although, several methods are used to reduce the parameters and resolve 'overparamterization' issues, there is huge scope to leverage the recent advances on the intersection between Graph Neural Networks(GNNs) and Transformers i.e. Graph Transformers Yun et al. [2019]. We propose devising techniques to exploit the salient features of Graph Transformers to apply and improve upon solutions which are dealt by multi-relation GCNs.

1 Introduction

Graph neural networks have been widely used for modelling structured data. However, most of the GNNs used are designed for homogeneous graphs. Homogeneous graphs have nodes and edges belong to the same type. Heterogeneous Graph Transformer (HGT) Hu et al. [2020] was introduced for modelling heterogeneous graphs. This is majorly performed by characterising the heterogeneous feature across every node- and edge-type dependent parameters. Though, there were multiple attempts in the past, HGT stands outperforms the existing attempts due to the following reasons:

- The meta relation between the target node, source node and edge is being used to consider both command and specific patterns of relationships.
- Neural architecture is built to incorporate heterogeneous type neighbour information replacing customised meta paths.
- Most of the methods do not take dynamic nature into consideration, while HGT uses relative temporal encoding to incorporate the dynamics of the graphs

The ideas of HGT is to aggregate source information from the source nodes which provides a representation for the target node. This is done in three ways: (a) Heterogeneous Mutual Attention, (b) Heterogeneous Message Passing and (c) Target Specific Aggregation.

The novelty of HGT model is that how its architecture is used to maximise parameter sharing while still the maintaining the characteristics of different relations. This is specifically done by parameterizing the weight matrices between the interactions of the target, source nodes and edges as show in Figure 1.

On the other hand, CompGCN is a novel convolutional framework which embeds both nodes and relations in a relational graph. It also makes use of a variety of entity-relation composition operations leveraged with techniques from Knowledge graph embedding and scales with number of relations.

The primary goal of this proposal is to leverage the novelty of attention based heterogeneous modelling to predict the CompGCNs tasks: (a) Link Prediction, (b) Node Classification and (c) Graph Classification.

$$\mathbf{Attention}_{HGT}(s,e,t) = \underset{\forall s \in N(t)}{\text{Softmax}}\Big(\underset{i\in[1,h]}{\Big\|} ATT\text{-}head^i(s,e,t)\Big) \tag{3}$$

$$ATT\text{-}head^i(s,e,t) = \Big(K^i(s)\, W^{ATT}_{\phi(e)}\, Q^i(t)^T\Big) \cdot \frac{\mu_{\langle \tau(s),\phi(e),\tau(t)\rangle}}{\sqrt{d}}$$

$$K^i(s) = \text{K-Linear}^i_{\tau(s)}\Big(H^{(l-1)}[s]\Big)$$

$$Q^i(t) = \text{Q-Linear}^i_{\tau(t)}\Big(H^{(l-1)}[t]\Big)$$

Figure 1: HGT Architecture

2 Datasets

The paper proposed to use attention based modelling for CompGCN. Hence, the following datasets can be used:

- Wikikg2 Dataset: It is a subset of Wikidata knowledge baseVrandečić and Krötzsch [2014], containing 2,500,604 entities and 535 relation types with 17,137,181 edges.
- WN18RR Dataset : is a link prediction dataset created from WN18, which is a subset of WordNetMiller et al. [1990]. WN18 consists of 18 relations and 40,943 entities. However, many text triples are obtained by inverting triples from the training set. Thus the WN18RR dataset is created to ensure that the evaluation dataset does not have inverse relation test leakage. WN18RR dataset contains 93,003 triples with 40,943 entities and 11 relation types.
- FB15k-237 Dataset : FB15K-237 is a variant of the Fb15K Freebase dataset where inverse relations are removed. Fb15K has a total of 592,213 triplets with 14,951 entities and 1,345 relationships. FB15K dataset suffered from major test leakage through inverse relations, where a large number of test triples could be obtained by inverting triples in the training set.

3 Solution Plan

The solution plan can be divided into the following:

- **CompGCN And HGT Implementation**: Execute the codes for datasets separetely for both the models.
- **Isolate salient features of HGT and seek ways to apply them to CompGCN**: Capture the novelty of HGT implementation in reducing the weight matrices of the parameters in different entity relationships.
- Integrate the attention based modelling into knowledge graphs to make it more efficient.
- Add rich-text metadata to make the Graph model even more data-rich (for example, using Wikipedia Data) and efficient for solving multi-relational problems.

4 Evaluation Plan

The evaluation plan is based on three main parameters on State of the Art Models such as InterHT, TripleRE and CompGCN itself:

- **Mean Reciprocal Rank(MRR)**:average of the reciprocal ranks of results for a sample of queries Q.
- **Mean Rank(MR):** average of the ranks of results for a sample of queries Q.
- **Hits Ratio at n(HR@n)**: a way of calculating how many "hits" you have in an n-sized list of ranked items.

5 Schedule

- Week 4: In this week, we plan to finish thorough literature survey and complete basic code for reading
- Week 5: In this week, we plan to finish thorough literature survey and complete basic code for reading different datasets.
- Week 6: Run end to end code on CompGCN and isolate salient features of HGT.
- Week 7: Study and model these salient features. Implement, run and train end to end with.
- Week 8: Study the results and start understanding Wikipedia dataset's API.
- Week 9: Figure out best way to infuse the Wikipedia data and implement code for the same.
- Week 10: Project Completion, accumulate and understand results and complete report.

6 Work Division

The work division will be done between the group members equally. PHD candidate Ziniu is guiding us and the distribution will be done as we move forward in the tasks. We all did the literature survey together and have a basic idea of what to do in the project.

References

Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of The Web Conference 2020*, pages 2704–2710, 2020.

George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4): 235–244, 1990.

Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. Composition-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1911.03082*, 2019.

Denny Vrandečić and Markus Krötzsch. Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. Graph transformer networks. *Advances in neural information processing systems*, 32, 2019.