CS 577: Natural Language Processing

Investigating Knowlege Graph Completion with Pre-trained Language model and Graph Neural Networks

Amit Roy
Taisuke Mori

Computer Science
Purdue University

Outline

- Introduction
- Methods
- Experimental Design
- Results
- Discussion

Introduction

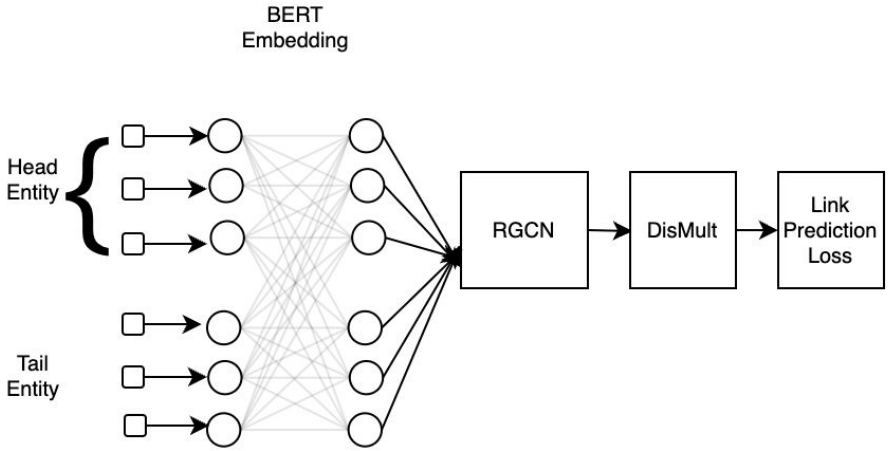
- Predicting missing links between entities in knowledge graph is useful for many applications
  - Question answering
  - Information retrieval, etc.

- Several possible approaches
  - GNN-based: encode graph structure well (RGCN)
  - Language model-based: exploit textual information related to entities and relations (BERT)

- We aim to combine two lines of research together to achieve the best of both worlds!

Method : BERT-RGCN

- Goal: learning a function that gives a higher score for a true triplet (entity1, relation, entity2)
- We feed pre-trained BERT embeddings to the input layer of R-GCN as entity and relation representation
- Due to memory constraint, we fix the BERT embeddings during training and test



Experimental Design

- Datasets
  - WN18, WN18RR, FB15k-237
- Baseline
  - R-GCN
- Evaluation
  - Create a ranking for each true triplet with dummy triplets based on the scores of the model and compute the ranking metrics over test set
    - MRR
    - Hits@k (k = 1, 3, 10)

Triplet	Score
(Purdue, is_in, TX)	80
(Purdue, is_in, IN)	75
(Purdue, is_in, IL)	70
(Purdue, is_in, CA)	60
(Purdue, is_in, NY)	50

Results: WN18

Dataset	R-GCN	BERT-RGCN
MRR (filtered)	**0.003353**	0.002122
Hits (filtered) @ 1	**0.000729**	0.000298
Hits (filtered) @ 3	**0.001943**	0.001043
Hits (filtered) @ 10	0.006412	**0.006553**

Results: WN18RR

Dataset	R-GCN	BERT-RGCN
MRR (filtered)	0.000228	**0.000693**
Hits (filtered) @ 1	0.000002	**0.000160**
Hits (filtered) @ 3	0.000005	**0.001043**
Hits (filtered) @ 10	0.000008	**0.000957**

Results: FB15k-237

Dataset	R-GCN	BERT-RGCN
MRR (filtered)	**0.000798**	0.000657
Hits (filtered) @ 1	**0.000400**	0.000024
Hits (filtered) @ 3	**0.000600**	0.000513
Hits (filtered) @ 10	**0.001200**	0.000684

Discussion

- For WN18RR, BERT-RGCN performs better than R-GCN
- However, the other two datasets, the performance of BERT-RGCN is worse
- Possible reason:
  - Training does not proceed well
  - Possibly the model underfits because embeddings are fixed
  - With fine-tuning, the performance may improve

Training Loss for WN18 dataset (100 - 1000 epochs)