Deep dive into Machine learning by leveraging Networks

WE Machine Learning Project

Aarushi Gulati, Samhitha Bharthulwar, Saumya Chaturvedi

WE Program Cohort 4

December 25, 2023

Building blocks - Triplets

"New Delhi is the capital of India."

(text)

head: India, relation: capital, tail: New Delhi

(triplet)

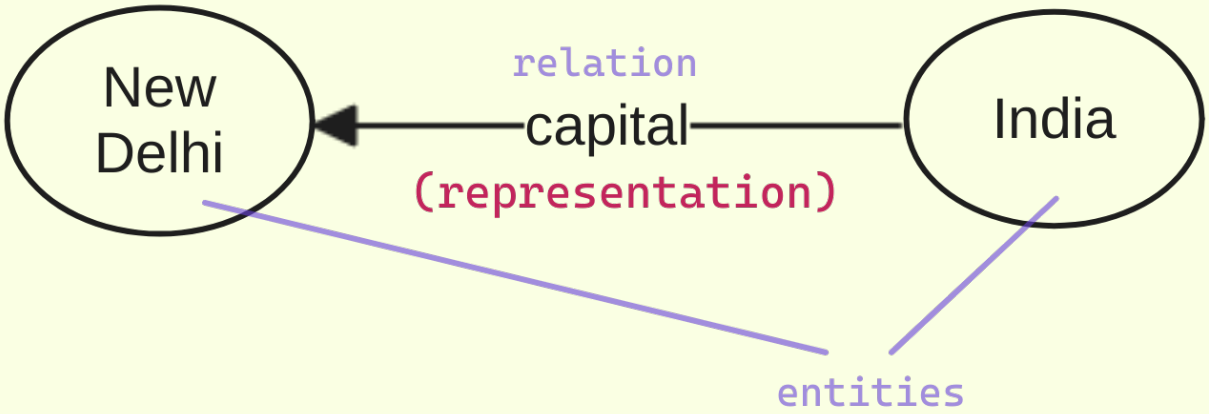

Figure 1: Building a triplet

Paper Review [1]

- The problem: representation of KG embeddings, no specific models
- Existing Translation-based and Semantic matching models
- considered only KG triplets, suffer from structure sparsity
- Text-enhanced representation: using CNNs and LSTMs

Paper Review [1] (cont.)

Proposed Method (Teger):

- Triplet Embedding: $f(h, r, t) = -||h + r - t||^2$
- Auxiliary text encoding: Text-graph construction: $G = \{V, E\}$
- Knowledge Graph Fusion: integrating auxiliary embeddings and triplet using gating vector = CONVOLUTION!
- End-to-end training = Loss = margin between incorrect (samples) and correct triplets

Progress with KG creation

Challenges with models we tested:

- Mistral-7B: Computationally expensive
- FRED: The links were ambiguous
- REBEL: Had a token limit of 1024 tokens (roughly 730 words)

Break down corpus into parts? Some information will be lost.
Solution (or rather work-around) would be to continue working with end-to-end models but with smaller input and parallely check out separate models for NER and RC.

Visualisation

- NetworkX and PyVis, don't support heterogenous graphs
- Graphviz, Neo4j and Pytorch Geometric support heterogenous graphs
- Neoj4 is interactive

Next Steps

- Test separate models for NER and RC instead of an end-to-end model
- Choose best combination(s) from multiple NER and RC models
- Link Prediction models and applying GNNs to KG processing tasks

References

[1] Linmei Hu et al. "Text-Graph Enhanced Knowledge Graph Representation Learning". In: *Frontiers in Artificial Intelligence* 4 (2021). ISSN: 2624-8212. DOI: 10.3389/frai.2021.697856. URL: https://www.frontiersin.org/articles/10.3389/frai.2021.697856.