

INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY, HYDERABAD

CS4.501 SOCIAL COMPUTING SPRING-2021

COURSE PROJECT:LITERATURE SURVEY

Enhance Knowledge Graphs/Knowledge bases like Wikidata

Team Members:

Pushpa Yadhav (2019900034)

Vijaya Lakshmi (2018900071)

Karthikeyan Arumugam(2018900074)

Mentor:

Tushar Abhishek

January 28, 2021



INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY

HYDERABAD

1 Introduction

Human knowledge provides a formal understanding of the world. Knowledge graphs that represent structural relations between entities have become an increasingly popular research direction towards cognition and human-level intelligence. Some of widely used Knowledge graphs in NLP related research are Freebase, YAGO and WikiData. In this project we would be exploring ways to make wikidata denser.

2 Problem statement

The amount of facts present in Knowledge graphs (KG) does not capture most of the world knowledge which is present in the text. The Quality of the Knowledge graph can be illustrated by density of edges (relations) and volume of information captured in it. There are various possibilities to enrich and improve the quality of the KGs.

Some of possible challenges at present are

- Knowledge graphs contains entities but not all relations are captured
- Entities are not captured in Knowledge graphs hence information pertaining to those entities are missing
- Augmenting different KGs/Sources to create enrich KG is tedious task.

Some important question in front of us is

- How to capture more information and enrich existing KGs?
- How to create more edges connecting entities which are already captured in Wikidata ?
- how to Ingrain more world knowledge into existing KGs Effectively ?
- is it possible to Augment different KGs and scattered information to enrich KGs such as wiki data. ?

3 Scope for this project work

- We will be using WikiData as our Base Knowledge Graph
- We will be focusing on Only one specific Domain for this project
- We will be using Wikipedia pages as source to enrich domain we choose

4 Literature Review

- **A Survey on Knowledge Graphs:Representation, Acquisition and Applications , Ji et al.,2021**

This research conducted a comprehensive survey on 4 scopes

1. knowledge graph embedding
2. knowledge acquisition of entity discovery
3. temporal knowledge graph representation learning
4. real-world knowledge-aware applications

This paper discussed deeper into theory and methods involves in various aspects of Knowledge graphs , they covered technical aspects such as knowledge representation, techniques to acquisition of new knowledge , various models and their comparison in detail. K-BERT,LSTM-CNN, RNN, BiLSTM,reinforcement learnings,and various other methods were discussed in detail and how it can be used in various stages of Knowledge graph lifecycle. This gave us lot of insight in state-of-art research effort going in Knowledge graph related area.

- **Accurate Text-Enhanced Knowledge Graph Representation Learning, Bo An et al., 2018**

This paper proposes representation frame work called "an accurate text enhanced knowledge graph representation framework", which enhance the knowledge representations of a triple, and effectively handle the ambiguity of relations and entities through a mutual attention model between relation mentions and entity descriptions. Their experiment results shows that their method can achieve the state-of-the-art performance, and significantly outperforms previous text-enhanced knowledge representation models.

- **Learning to Update Knowledge Graphs from Reading News, (Tang et al., 2019)** This research work proposes a novel graph based neural network method called GUpdater. GUpdater is build upon graph neural network (GNN) with a text based attention model. This model was able to effectively perform link-adding or link-deleting operations to ensure the KG up-to-date according to news snippets. Experiments demonstrated that this model can handle explicit and implicit information found in news sources.

- **Collective Multi-type Entity Alignment Between Knowledge Graphs, Qi Zhu1 et al.,2020**

This research paper presents a new method "Collective Graph neural network for multi-type entity alignment" (CG-MuAlign). This method jointly aligns multiple types of entities, collectively leverages neighborhood information and generalizes to unlabeled entite types. This experiment propose

novel collective aggregation function tailored for reliving the incompleteness of the knowledge graphs via both cross-graph and self attentions, it also scales up effectively with mini-batch training paradigm and effective neighborhood sampling strategy. Their experiments with real world knowledge graphs with millions of entities and they observed superior performance beyond existing methods. Running time of this method is much less than current state-of-the-art deep learning methods. Experiments demonstrated that this proposed method can handle multiple knowledge graphs alignment simultaneously.

- **Language Models are Open Knowledge Graphs, Chenguang Wang et al., 2020**

This paper proposes an unsupervised method to cast the knowledge contained within language models into KGs. Specifically, it shows how to construct knowledge graphs (KGs) from pre-trained language models (e.g., BERT, GPT-2/3), without human supervision. This paper introduces a two-stage unsupervised approach called MaMa (Match and Mapping), which can successfully recover the factual knowledge stored in language models to build KGs from scratch. This MaMa constructs a KG with a single forward pass of pretrained language models over a textual corpus. Further experiments with this model demonstrated that open knowledge graph features new facts when compared to WikiData and TAC KBP. This model establishes a bridge between the deep learning and knowledge graphs, its results suggest that larger language models store richer knowledge than existing knowledge graphs.

5 Proposed Project work

By exploring above literature, we gained some insights on different approaches to enrich knowledge graphs, by adding more facts, increase the density of the relations established in the knowledge graph. After careful consideration we are inclined towards trying approaches mentioned in "Language models are open knowledge graphs" by Chenguang Wang et al., 2020.

1. In our project we are planning to use WikiData as Knowledge Graph and corresponding Wikipedia text as input to the language model.
2. Our end outcome of this project is to find how this method can enrich WikiData and analyse issues in the proposed model.
3. We would like to identify any issues and explore possible solutions for such shortcomings of this model.
4. With help of our experiments we would investigate and report the accuracy measures.

6 References

- A Survey on Knowledge Graphs: Representation, Acquisition and Applications , Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, Philip S. Yu, 2020 , arXiv:2002.00388 [cs.CL]
- "Accurate Text-Enhanced Knowledge Graph Representation Learning", Bo An, Bo Chen, Xianpei Han and Le Sun, 2018, "Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 , <https://www.aclweb.org/anthology/N18-1068>, 10.18653/v1/N18-1068
- Learning to Update Knowledge Graphs from Reading News, Jizhi Tang, Yansong Feng, Dongyan Zhao , Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) <https://www.aclweb.org/anthology/D19-1265>, 10.18653/v1/D19-1265
- Collective Multi-type Entity Alignment Between Knowledge Graphs, Qi Zhu , Hao Wei, Bunyamin Sisman, Da Zheng , Christos Faloutsos, Xin Luna Dong , Jiawei Han, 2020, Proceedings of The Web Conference 2020, <https://assets.amazon.science/ff/7a/b96282984a0fbe5e31a8fcf68d17/scipub-1202.pdf>, <https://doi.org/10.1145/3366423.3380289>
- Language Models are Open Knowledge Graphs , Chenguang Wang, Xiao Liu, Dawn Song , 2020, [arXiv:2010.11967](https://arxiv.org/abs/2010.11967)