Predict Missing Links Within a Knowledge Graph

A Project Report Submitted
in Partial Fulfillment of the Requirements
for the program of

Power Internship

By

Prataparao Sai Vamsi
Akashdeep Soni
Abhishek M

Under the Supervision of
Mr. Gagan Gayari
(Specialist Programmer)

Infosys Limited, Mysore
May, 2022

DECLARATION

We hereby declare that the work reported in the project report entitled as "**Predict Missing Links Within a Knowledge Graph**", in partial fulfillment for the completion of internship submitted at Infosys Limited, as per best of our knowledge and belief there is no infringement of intellectual property right and copyright. In case of any violation, we will solely be responsible.

Prataparao Sai Vamsi
Akashdeep Soni
Abhishek M

Power Interns
Infosys Limited.

CERTIFICATE

This is to certify that the work titled **"Predict Missing Links Within a Knowledge Graph"**, submitted by "**Prataparao Sai Vamsi, Akashdeep Soni, Abhishek M**" in partial fulfilment for the completion of internship at Infosys Limited. has been carried under my supervision. As per best of my knowledge and belief there is no infringement of intellectual property right and copyright. Also this work has not been submitted partially or whole to any other organizations. In case of any violation concern student will solely be responsible

Signature of Supervisor

29th MAY 2022

ACKNOWLEDGEMENT

Any endeavor cannot lead to success unless and until a proper platform is provided for the same. This is the reason we find ourselves very fortunate to have undergone our major project work under the supervision of **Gagan Gayari.**

Our sincere gratitude to **Gagan Gayari**, our project supervisor for having faith in us and thus allowing us to carry out a project on a technology completely new to us. He helped immensely by guiding us throughout the course of the project.

Prataparao Sai Vamsi
Akashdeep Soni
Abhishek M
29th MAY 2022

ABSTRACT

Link Prediction is the problem of predicting edges that either don't yet exist at the given time t or exist, but have not been discovered, are likely to occur in the near future. We develop approaches to link prediction based on measures for analysing the proximity of nodes in a network. Consider a co-authorship network among scientists, e.g., two scientists who are close in the network will have colleagues in common, so they are more likely to collaborate in the near future. Our goal is to make this intuitive notion precise and to understand which measures of proximity in a network lead to the most accurate link predictions.

TABLE OF CONTENT

INTRODUCTION

A Knowledge Graph (KG) is simply a heterogenous graph with nodes and edges capturing some semantic information about real world entities and concepts.

For example, A person and a city can be the nodes and lives_in is the edge connecting them. The triple format representation for the network is then (person, lives_in, city).

PROBLEM STATEMENT

Given a Knowledge Graph, predict the possible links that could appear among nodes.

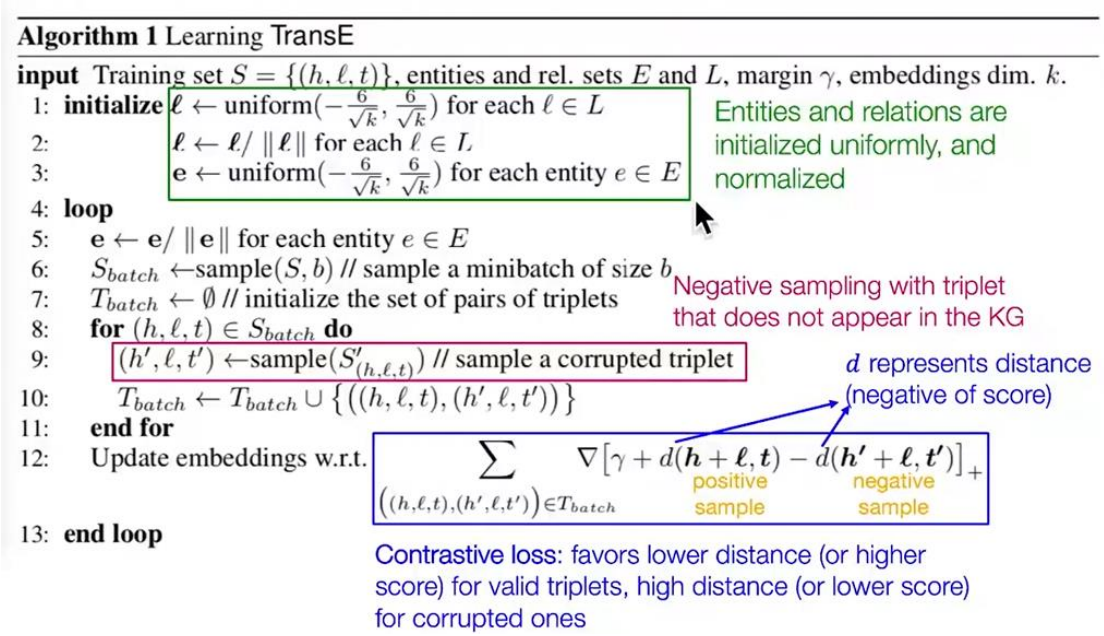
E.g. lf we have a relation (Varun, born_in, Pune) within the Knowledge Graph, then based on the semantics and information present within that knowledge graph there should exist some relation like (Varun, citizen_of, India), which is not already present in the graph.

OBJECTIVE

Our objective is to propose a canonical model which is easy to train, contains a reduced number of parameters and can scale up to very large databases. Hence, we propose TransE, a method which models relationships by interpreting them as translations operating on the low-dimensional embeddings of the entities. Despite its simplicity, this assumption proves to be powerful since extensive experiments show that TransE significantly outperforms state-of-the-art methods in link prediction on two knowledge bases

METHODOLOGY

1. Loading a KG and creating train/test splits
2. Training and evaluating a KGE Model
3. Testing user hypothesis
4. Early stopping and types of evaluation
5. Choosing model hyperparameters
6. Discovering facts using trained model
7. Visualizing embeddings and Clustering



8. The basic idea behind our model is that we want that h + l ≈ t when (h, l, t) holds (t should be a nearest neighbor of h + l), while h + l should be far away from t otherwise.

DATASET

FB15K-237

This dataset is a variant of the original dataset where inverse relations are removed, since it was found that a large number of test triplets could be obtained by inverting triplets in the training set. Total triplets are 310079.

LIBRARY

Ampligraph

AmpliGraph is a suite of neural machine learning models for relational Learning, a branch of machine learning that deals with supervised learning on knowledge graphs. Its an open source Python library that predicts links between concepts in a knowledge graph.

RESULT AND CONCLUSION

The objective of the project is to predict the missing parts in the graph. We have successfylly completed the project with high and efficiency. This projects helps to achieve link prediction. Link prediction works with subject correction and object correction for better performance.

FEASIBILITY STUDY

Technical Feasibility

Project is technically feasible as all the technical requirements have been analysed and are easily obtainable.

Operational Feasibility

Project is operationally feasible as the trained model has been saved which can restored in future.

Economic Feasibility

Project is economically feasible as it will be build using open source softwares and library which are free to use.

Legal Feasibility

Project is legally feasible as all modules used permit usage from open source and non-monetized applications.

Schedule Feasibility

Project is schedule feasible as the project can be completed within the said deadline.

FUTURE SCOPE

1. Link prediction helps to understand associations between nodes in social communities. We can infer new interactions among its members which are likely to occur in the near future

2. Identifying the structure of a criminal network by predicting missing links in a criminal network using incomplete data.

3. Build recommendation systems (e-commerce)

References

1. **Stanford CS224W: Machine Learning With Graphs**
(https://www.youtube.com/watch?v=JAB_plj2rbA&list=PLoROMvodv4rPLKxIpqhjhPgdQy7imNkDn&ab_channel=StanfordOnline)

2. **Knowledge Graph Embedding Tutorial**
(https://youtu.be/gX_KHaU8ChI)

3. **Ampligraph Docs**
(https://docs.ampligraph.org/en/1.4.0/)