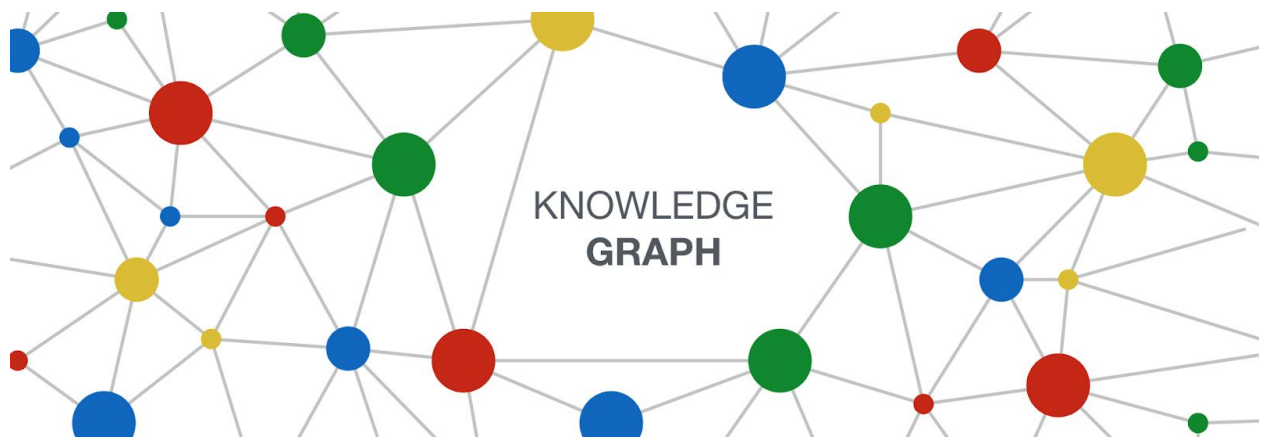


Knowledge Graph Study Aid



Omar Sanchez
osanchez@umass.edu

Cyrus Freshman
cfreshman@umass.edu

Vidul Mahendru
vmahendru@umass.edu

1. Introduction	3
2. Related Work	3
3. Your Approach	4
3.1 Milestones and Schedule	4
4. Data	5
5. Tools	5
6. References	5

1. Introduction

The focus of our project will be to create a knowledge graph for a specific context that can be given words or phrases as queries and return relevant information in a synthesis of question answering and summarization. Our goal is to allow users to quickly search any given collection of text for a specific concept and receive a summarization of that concept and its links to other concepts in the context of the given text. The inspiration for this project comes from being full time college students. As college students we are required to read large and complex pieces of text and digest this information in small time frames. Learning something for the first time is not always easy, so with this project we hope to create a tool that not only students, but anyone interested in learning, can use to make learning more efficient. Uses for our program can also be expanded out to include medical journals or books so that people have a resource for diagnosing or treating themselves of disease.

2. Related Work

We came across a paper where a group of researchers aimed to find a solution to creating a “complete” knowledge graph. Knowledge graph completion aims to predict relations between entities under supervision of the existing knowledge graph. Knowledge graph completion is useful for finding new relational facts, an important supplement to relation extraction from plain texts. According to the paper, Translating on embeddings (TransE) and Translating on Hyperplanes (TransH) are models currently used in creating knowledge graphs. These models build entity and relation embeddings by regarding a relation as translation from head entity to tail entity. Both of these models place entities and relations within the same semantic space, but an entity may have various relations and aspects which make these graphs insufficient for modeling. This is where a new method of modeling called TransR is introduced, which the researchers claimed could be used to build entity and relation embeddings in separate entity space and relation spaces.

Pros:

The benefits of using TransR for modeling include its ability to model entities and relations in distinct spaces, and perform translation in the corresponding relation space. TransR showed similar efficiency to TransH, but showed significant improvement overall when compared to TransE.

Cons:

Knowledge graph completion is similar to link prediction in social network analysis, which is used to create recommendations to its users (recommended friends, products, and destinations; data mining) but is more challenging for the reasons listed: “(1) nodes in knowledge graphs are entities with different types and attributes; and (2) edges in knowledge graphs are relations of different types. For knowledge graph completion, we not only determine whether there is a relation between two entities or not, but also predict the specific type of the relation” (Lin et al. 2015).

In another paper, researchers used variational reasoning to create “efficient” queries for a knowledge graph. According to these researchers, “when people ask questions, their expressions are noisy (for example, typos in texts, or variations in pronunciations), which is nontrivial for the QA system to match those mentioned entities to the knowledge graph” (Zhang et al 2017). They also noted that “many questions require multi-hop logic reasoning over the knowledge graph to retrieve the answers”(Zhang et al 2017). In order to solve both of these issues they proposed a unified deep learning architecture and an end-to-end variational learning algorithm which can handle noise in questions, and learn multi-hop reasoning simultaneously. The researchers were able to achieve improved results to previous benchmarks. In their finding they found that TransH overcomes many of the flaws of TransE concerning the reflexive/one-to-many/many-to-one/many-to-many relations while inheriting its efficiency.

Pros:

The benefits of using TransH over TransE is that it is more efficient. TransH aims to create a good trade-off between model complexity and efficiency so that it can overcome the flaws of TransE while inheriting its efficiency.

Cons:

TransE has many flaws when dealing with relations with mapping properties of reflexive/one-to-many/many-to-one/many-to-many. Some advanced models with more free parameters are capable of preserving these mapping properties, however, the model complexity and running time is significantly increased accordingly. As a cost to preserving these properties, predictive performances of the advanced models are even worse than TransE.

3. Your Approach

Our team plans on creating a knowledge graph using knowledge graph embeddings with Translation on Hyperplanes (TransH). This method delivers similar efficiency to TransE, but allows us to work with different mapping properties that we believe will be useful for our goal. We will create our knowledge graphs using a similar manner proposed in the papers. We will then devise a system for querying this knowledge graph for a specific concept and returning a summary of the information pertaining to this concept, consisting of its main attributes and links to other concepts.

3.1 Milestones and Schedule

All parts of the project will be worked on as a group. We will meet at least weekly to work on each of the milestones. The main source of communication will be Slack, and project documents and code will be managed on GitHub.

1. Acquire and preprocess data 10/20 – 10/29 (10 days)
2. Build models for task 10/30 – 11/12 (14 days)
3. Write progress report part 11/13 – 11/16 (4 days, due 11/16)

4. Analyze the output of the model, do an error analysis 11/17 – 11/30 (14 days)
5. Work on final report and presentation 12/1 – 12/13 (14 days)

4. Data

There are a number of question answering datasets that we will be able to use to train and evaluate our with model such as Facebook's [bAbI dataset](#) and Stanford's [SQuAD 2.0](#), which consist of questions and answers based on context text. We will initialize our word embeddings with pre-trained word vectors such as those available from [GloVe](#).

5. Tools

We plan to use the TensorFlow library for deep learning. We will likely need to use Google's cloud GPU/TPU services for training our model.

6. References

Lin, Yankai, et al. "Learning entity and relation embeddings for knowledge graph completion." *AAAI*. Vol. 15. 2015.

Zhang, Yuyu, et al. "Variational reasoning for question answering with knowledge graph." *arXiv preprint arXiv:1709.04071* (2017).