Leveraging BERT with Knowledge Graph Embedding for Simple Question Answers
Experimental Protocol

Ethan Nguyen, Vu Pham, Mohan Rangarajan

April 15, 2020

1. Hypothesis

Simple Question Answering (QA) using Knowledge Graph (KG) utilizes the facts in KG to answer questions posed in natural language. For a simple question i.e. a question that has only one head entity and a single predicate in the KG, the associated tail entity provides the answer. For retrieving a response, KG can be queried directly or response can be retrieved from the KG embedding space. KG embedding approaches use a low dimensional vector to represent the entity and predicate thereby preserving their relationship. Critical to identifying the fact in a KG associated with a natural language question is identifying the predicate and unambiguously identifying the head entity in the question.

Our hypothesis pertains to simple questions. We hypothesize that, utilizing variants of pre-trained Bidirectional Encoding Representation for Transformers (BERT)[1] , in conjunction with a fine-tuned model will help to better disambiguate the head entity, uniquely identify the predicate and yield a set of candidate entity tokens for narrowing down facts in the KG. This will result in improved accuracy of KG based QA systems.

2. Datasets

For our experiment, we plan to use FB2M and FB5M which are subsets of the publicly available Freebase dataset[2]. We leverage the pre-processing performed by Xiao Huang et. al[3] on the Freebase dataset. The pre-processing includes deleting repeated facts, utilizing entity name collection to build mapping between entities and their names. We plan to use the training, validation and test splits provided in SimpleQuestions [4].

3. Metrics

In the SimpleQuestions [4] dataset, each question has only one correct answer (and no blanks or multiple answers one of which is incorrect). As a result, we won't have any True Negatives (i.e. question does not have an answer and model also predicts no answer) or False Positives (i.e. question has a wrong answer and we fail to recognize the wrong answer). Consequently, we will use "Accuracy" in predicting the answer based on the (head entity, predicate) tuple as the evaluation criterion. This criterion allows us to measure the performance of the model and is also consistent with the metric used by Xiao Huang et. al[3] (our baseline). In general, we plan to use the methodology adopted by Xiao Huang et. al[3] for our project.

4. Models

Xiao Huang et. al[3] use a total of 3 different models –

1. Bidirectional LSTM with Attention for learning the predicate
2. Bidirectional LSTM with Attention for learning head entity.

3. A separate Bidirectional LSTM (without the Attention layer) with a fully connected layer and Softmax layer for Head Entity Detection. The goal of the Head Entity Detection model is to identify a set of candidate tokens as the name of the head entity. This set of tokens will help narrow down the search space in the KG embedding to entities with similar names. The role of model #2 is to reduce ambiguity in the search space.

For the KG embedding, we plan to use an implementation of low dimensional representation using TransE, a translation-based model proposed by Bordes et. al [5]. We will use the combined outcome of the 3 models as our baseline. As part of our efforts, we will replace the 3 models used with variants of pre-trained BERT models and fine tune those models to examine the effect on accuracy and further explore our hypothesis.

In addition we plan to also apply pre-trained BERT variants to the models in Salman Mohammed et. al [6].

5. General Reasoning

We plan to use the Freebase FB2M, FB5M datasets as the content for the KG. For a low dimensional representation of the KG, we plan to use TransE proposed by Bordes et. al [5]. This embedding is already available to us as part of the work done by Xiao Huang et. al [3]. We plan to replace the models developed by Xiao Huang et. al [3] with pre-trained BERT model variants with appropriate fine tuning to learn the head entity (entity learning model), the predicate (predicate learning model) and a set of candidate head entity names (head entity detection model). The candidate head entity names will be used to narrow down the entities in the KG embedding space. We will use the SimpleQuestions [4] dataset to train and evaluate the models. We plan to use the training, validation and test splits provided in SimpleQuestions [4]. The tuple of predicted head entity, predicted predicate will be used to predict the answer. The answer is calculated based on the facts in the KG embedding that is closest to the predicted tuple. We will use "Accuracy" in predicting the answers as the evaluation criterion. This criterion allows us to measure the performance of the model and is also consistent with the metric used by Xiao Huang et. al [3] (our baseline).

6. Summary of Progress So Far

We have downloaded the models developed by Xiao Huang et. al [3] and the associated Freebase dataset and the SimpleQuestions dataset. We have compiled these in preparation for confirming the baseline scores reported. In addition, we have also downloaded the KG embedding implementation by Bordes et. al [5]. We have replicated the test results reported by Xiao Huang et. al [3] . We are exploring different variations of BERT for use with learning the head entity, predicate and head entity detection. We have started to focus on training the head entity detection model using BERT and evaluating the results of this model.

7. References

[1] J. D. M.-W. C. K. L. K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *arXiv:1810.04805*, 24 May, 2019.

[2] Google, "Freebase Data Dumps," 2018. [Online]. Available: https://developers.google.com/freebase. [Accessed April 2020].

[3] J. Z. D. L. P. L. Xiao Huang, "Knowledge Graph Embedding Based Question Answering," in *The Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)*, Melbourne, VIC, Australia, 2019.

[4] N. U. S. C. a. J. W. Antoine Bordes, "Large- Scale Simple Question Answering with Memory Networks.," in *arXiv preprint arXiv:1506.02075*, 2015.

[5] N. U. A. G.-D. J. W. a. O. Y. Antoine Bordes, "Translating Embeddings for Modeling Multi-relational Data," pp. 2787-2795, 2013.

[6] P. S. J. L. Salman Mohammed, "Strong Baselines for Simple Question Answering over Knowledge Graphs with and without Neural Networks," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018.

[7] R. S. a. C. M. Jeffrey Pennington, *EMNLP,* p. 1532–1543., 2014.