Learning Knowledge Graph Paths with BERT

Proposed by Jason Youn

# Overall Aim

The goal of this project is to develop a novel approach for training a language representation model from a knowledge graph (KG) to produce state-of-the-art results for a wide range of KG-related tasks, such as triple classification, link prediction, and relation prediction.

# Background and Challenges

Recently, many data are reported using large-scale KGs (*e.g.* FreeBase1, YAGO2, and WordNet3) due to their advantages in identifying patterns among data and utilizing the massive information content they carry to gain insights into the mechanisms of action, associations, and testable hypotheses4,5. However, since KGs are often incomplete6, a fundamental problem is to predict relations between the entities based on existing triples in the KG. Recently, KG-BERT7 represented triples as text sequences and used BERT8, a language representation model utilizing the attention mechanisms9,10, to learn scoring function for relation prediction. BERTRL11 further increased the performance by incorporating local path information when fine-tuning BERT. However, both methods simply use the pre-trained BERT trained on free-text like BooksCorpus12 and English Wikipedia. There still is room for improvement by pre-training BERT with KG and learning from the paths in the KG.

# Significance and Impact

By developing a full training scheme (both pre-training and fine-tuning) for BERT which utilizes the path information in the KG, we hypothesize that the performance of the KG-related downstream tasks will improve. Moreover, the proposed method will be an explainable model in which users gain insight into why a prediction was made by analyzing the paths used for inference. We also expect that the training speed will be faster with far fewer parameters since the input to BERT will be significantly smaller (512 tokens in original BERT as opposed to <128 expected tokens in our approach).

# Relevance to Other Projects

**Relevance to KIDS:** The proposed system can be swapped into the KIDS to serve as a hypothesis generator (HG). We hypothesize that the proposed system provides an improved hit rate compared to the original PRA-MLP-based HG.

**Relevance to AIFS:** The proposed system can be applied to the automatically curated knowledge base of food and chemical compound information for making an explainable predictor of health effects from food chemical composition.

# Approach and Success Metric

**Specific aim 1. Develop a novel BERT pre-training process given a KG as input.**

**Approach:** The system first generates paths of bounded lengths from the KG (**Figure 1**). We then feed pairs of paths as input for pre-training the BERT. If these two paths can be connected to form a single valid path, we predict True for the Valid Path Prediction (VPP) task, and False otherwise. Masked Language Model (MLM) task is the same as the original BERT implementation (**Figure 2**). Also, we propose using the type embeddings in addition to the segment and position embeddings (**Figure 4**) which denote the type of each token. Note there exists two separate type embeddings for the forward and inverse relation.

**Success metrics:** The training time of original BERT trained on the document-level corpus (*e.g.* Wikipedia, BooksCorpus, etc.) will be compared with that of our KG-based proposed system. Also, the number of parameters in the model will be compared to BERTBASE and BERTLARGE.

**Specific aim 2. Fine-tune the KG pre-trained BERT for various downstream KG-related tasks.**

**Approach:** Using the KG pre-trained BERT from the previous step, fine-tune the proposed system for triple classification, link prediction, and relation prediction tasks (**Figure 3** shows the link prediction as an example). For fine-tuning, we use the paths identified by PRA that connect the target entities for reasoning the question triple. This is similar to the question-answering task in the original BERT implementation.

**Success metrics:** Models will be trained and tested on common benchmark datasets for a fair comparison with other state-of-the-art models. Triple classification will use the datasets WN11 and FB13 with accuracy used for reporting the results. Link prediction will use the datasets WN18RR, FB15k-237, and UMLS with Mean Rank (MR) and Hits@10 used for reporting the results. Finally, relation prediction will use the dataset FB15k with MR and Hits@1 used for reporting the results.

**Specific aim 3. Application to the KIDS and AIFS project.**

**Approach:** We will apply the proposed system to the KIDS and AIFS project and compare it with other methods. For the KIDS, the proposed system will act as a hypothesis generator for generating higher hit-rate hypotheses. For the AIFS project, the proposed system will be used as a predictor of health effects from food chemical composition.

**Success metrics:** Existing metrics from each project will be used to analyze the performance benefit of the proposed system.

# Deliverables

The deliverables will be a software package written with Python, Tensor2Tensor library15, PyTorch (or TensorFlow). All source code will be deposited in Github with proper documentation for public access. Optionally, an interactive website or docker will be provided for the interested user to play with their own data.

# Timeline

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Apr | | May | | | | June | | | | July | | | | August | | | | Sep | |
| **Aim 1.** | Code study & brainstorming |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Development |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Evaluation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Documentation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Aim 2.** | Development |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Evaluation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Documentation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Manuscript |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Aim 3.** | Development |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Evaluation |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# References

1. Bollacker, K., Evans, C., Paritosh, P., Sturge, T. & Taylor, J. Freebase: a collaboratively created graph database for structuring human knowledge. in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* 1247–1250 (2008).

2. Suchanek, F. M., Kasneci, G. & Weikum, G. Yago: a core of semantic knowledge. in *Proceedings of the 16th international conference on World Wide Web* 697–706 (2007).

3. Miller, G. A. WordNet: a lexical database for English. *Commun. ACM* **38**, 39–41 (1995).

4. Li, Y. & Chen, L. Big biological data: challenges and opportunities. *Genomics. Proteomics Bioinformatics* **12**, 187 (2014).

5. Silvescu, A., Caragea, D. & Atramentov, A. Graph databases. *Artif. Intell. Res. Lab. Dep. Comput. Sci. Iowa State Univ.* (2012).

6. West, R. *et al.* Knowledge base completion via search-based question answering. in *Proceedings of the 23rd international conference on World wide web* 515–526 (2014).

7. Yao, L., Mao, C. & Luo, Y. KG-BERT: BERT for knowledge graph completion. *arXiv Prepr. arXiv1909.03193* (2019).

8. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv Prepr. arXiv1810.04805* (2018).

9. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv Prepr. arXiv1409.0473* (2014).

10. Luong, M.-T., Pham, H. & Manning, C. D. Effective approaches to attention-based neural machine translation. *arXiv Prepr. arXiv1508.04025* (2015).

11. Zha, H., Chen, Z. & Yan, X. Inductive Relation Prediction by BERT. *arXiv Prepr. arXiv2103.07102* (2021).

12. Zhu, Y. *et al.* Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. in *Proceedings of the IEEE international conference on computer vision* 19–27 (2015).

13. Veličković, P. *et al.* Graph attention networks. *arXiv Prepr. arXiv1710.10903* (2017).

14. Wang, G., Ying, R., Huang, J. & Leskovec, J. Direct Multi-hop Attention based Graph Neural Network. *arXiv Prepr. arXiv2009.14332* (2020).

15. Vaswani, A. *et al.* Tensor2tensor for neural machine translation. *arXiv Prepr. arXiv1803.07416* (2018).

# Figures

Diagram, schematic

Description automatically generated

**Figure 1. Sample knowledge graph used for demonstration.** The objective of the knowledge graph completion task is to find a possible relationship (*e.g.* lives\_in) between the two target entities (*e.g.* Ilias and Davis, CA).

Diagram

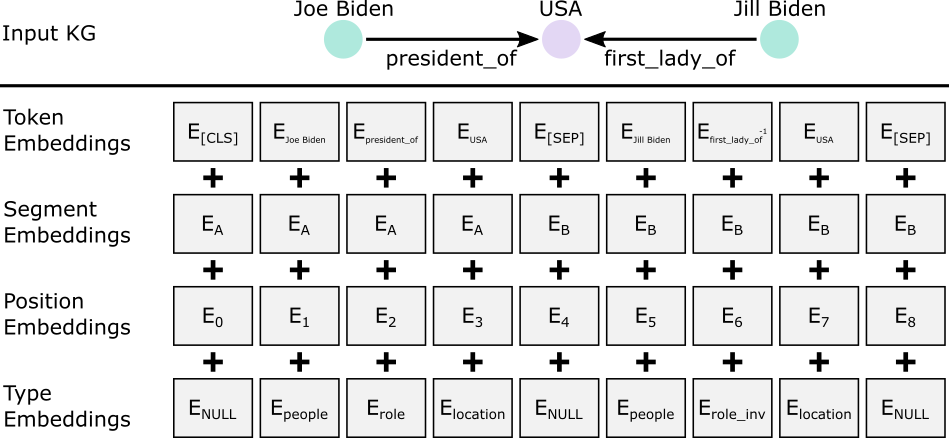
Description automatically generated

**Figure 2. Proposed pre-training scheme.**

A picture containing timeline

Description automatically generated

**Figure 3. Proposed fine-tuning scheme.**



**Figure 4.Input representation for the proposed system.**