Generating Knowledge Graphs from Large Language Models

1 Proposal

Over last ten years, significant progress has been made on large language models for a wide variety of natural language processing tasks (Min et al., 2021). While these models can achieve (and even occasionally supersede) human level performance, they are significantly more internally complex than historical approaches. Simple explanations of what each model has learned make interpretation difficult, and as these models are expected to be applied against an increasingly wide variety of tasks, having the ability to easily query their internal logic is increasingly important. Some of this can be achieved through the use of a knowledge graph (KG), which encodes semantic information and provides a way to visualize contextual relationships 1. Unfortunately, KGs have historically been complicated and onerous to create, requiring substantial manual input. A novel method to automatically generate KGs from a variety of large language models has been proposed by (Hao et al., 2022).

We propose comparing KGs generated before and after a fine tuning exercise to help understand what is being learned by the model. Utilizing a domain-specific language corpus during fine tuning should help narrow the scope of changes. The Geological Survey of Canada has provided a series of notebooks to generate a machine readable corpus from their existing catalog of publications (Raimondo and Lawley, 2022). We can then focus on a manual review of the major differences between the two KGs (before and after training) for geoscience specific words and concepts.
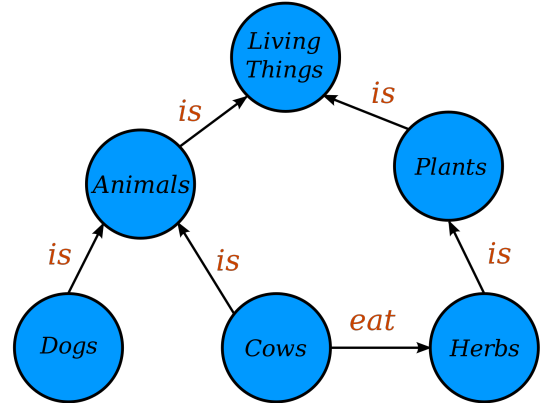

Figure 1: Example knowledge graph. [1]

References

Hao, S., Tan, B., Tang, K., Ni, B., Zhang, H., Xing, E. P., and Hu, Z. (2022). Bertnet: Harvesting knowledge graphs from pretrained language models.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., Agirre, E., Heinz, I., and Roth, D. (2021). Recent advances in natural language processing via large pretrained language models: A survey.

Raimondo, S., C. T. Z. A. B. L. K. D. H. J. B. S. N. G. and Lawley, C. (2022). Datasets to support geoscience language models.