Indicative Written Research Plan

Jiachen Li

1068299

What your research question is:

Utilising the pre-trained language model for named entity recognition to improve entity alignment performance in natural sentence generation from the knowledge graph.

How you plan to investigate the research question :

Trisedyal et al. proposed the GTR-LSTM model for natural sentence generation from a knowledge graph [1]. In entity alignment segmentation of the model, Trisedyal tested and suggested N-gram compositional function outperformance LSTM and SUM compositional functions [2]. However, after Trisedyal wrote the paper, pre-trained language representations like BERT have been proven to perform exceptionally well and quickly adapted to all NLP-related tasks to replace traditional methods. This paper plans to swap and update the compositional functions in the entity alignment part of the natural language generation (NLG) from N-gram to a pre-trained language model. And eventually, improve the final NLG performance.

How you will analyse the results of your investigation:

The evaluation will be done on three different levels. Firstly, benchmark pre-trained language with the baseline N-gram model in the named entity recognition [3]. Secondly, use the benchmark introduced by Zhang [4] to evaluate the performance of the proposed model in the entity alignment task. And lastly, use BLEU [5], METEOR [6], and TER [7]scores to evaluate the proposed model in NLG against Trisedyal's model.

Reference:

[1] B. D. Trisedya, J. Qi, R. Zhang, and W. Wang, "GTR-LSTM: A Triple Encoder for Sentence Generation from RDF Data," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Melbourne, Australia, 2018, pp. 1627–1637. doi: 10.18653/v1/P18-1151.

[2] B. D. Trisedya, J. Qi, and R. Zhang, “Entity Alignment between Knowledge Graphs Using Attribute Embeddings,” *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, Art. no. 01, Jul. 2019, doi: 10.1609/aaai.v33i01.3301297.
[3] A. Ahmed, A. Abbasi, and C. Eickhoff, “Benchmarking Modern Named Entity Recognition Techniques for Free-text Health Record De-identification.” arXiv, Mar. 24, 2021. doi: 10.48550/arXiv.2103.13546.
[4] R. Zhang, B. D. Trisedy, M. Li, Y. Jiang, and J. Qi, “A Benchmark and Comprehensive Survey on Knowledge Graph Entity Alignment via Representation Learning.” arXiv, May 05, 2022. doi: 10.48550/arXiv.2103.15059.
[5] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “Bleu: a Method for Automatic Evaluation of Machine Translation,” in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, Pennsylvania, USA, 2002, pp. 311–318. doi: 10.3115/1073083.1073135.
[6] M. Denkowski and A. Lavie, “Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems,” in *Proceedings of the Sixth Workshop on Statistical Machine Translation*, Edinburgh, Scotland, 2011, pp. 85–91. Accessed: Sep. 25, 2022. [Online]. Available: https://aclanthology.org/W11-2107
[7] M. Snover, B. Dorr, R. Schwartz, L. Micciulla, and J. Makhoul, “A Study of Translation Edit Rate with Targeted Human Annotation,” in *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, Cambridge, Massachusetts, USA, Aug. 2006, pp. 223–231. Accessed: Sep. 25, 2022. [Online]. Available: https://aclanthology.org/2006.amta-papers.25