The 6th International Conference on Interdisciplinary research on Computer science, Psychology, and Education (ICICPE' 2022)
December 26-28, 2022. Pattaya, Thailand.

Korean Commonsense Knowledge Graph based on GPT-3

Jaewook Lee[1] , Yuna Hur[2*]

[1] Korea University, Computer Science,
[2] Human-inspired AI Research
{jaewook133, yj72722}@korea.ac.kr

Abstract. Commonsense knowledge graph is a structured knowledge representation method that graphically expresses commonsense included in the corpus. The most widely used commonsense knowledge graph at present is ATOMIC. However, research to build this in Korean is not yet active. Therefore, in this work, we propose a method for constructing a Korean commonsense knowledge graph using existing commonsense-based datasets as inputs to one of the most powerful generative models, GPT-3.

Keywords: Commonsense, Knowledge Graph, Korean, GPT-3

1 Introduction

Commonsense inference is one of the major research fields in deep learning-based natural language processing. For natural commonsense inference, this knowledge must be reflected in the model. To this end, research has been actively conducted to express the reasoning process with a commonsense knowledge graph.

Representative studies include ATOMIC [1], which expresses a knowledge graph as an Event and If-Then relationship, deduces an ungiven relationship based on it, and Symbolic Knowledge Distinction [2], which automatically generates commonsense information by learning the results of a large-scale corpus-trained teacher model.

However, there is no commonsense knowledge graph composed of Korean. Therefore, this paper proposes a method of constructing a Korean commonsense knowledge graph using the KommonGen [3] dataset, which generates Korean sentences that conform to commonsense based on keywords, as input sentences for GPT-3 [4].

2 Method

In this paper, we generate a new inference sentence that considers the given premise sentence and relation tag by inputting the sentence of the KommonGen

* Corresponding Author

dataset and the relation tag of ATOMIC into the GPT-3 generation model. When the GPT-3 sentence generation model receives a prompt, the model infers and generates the content after the corresponding sentence. In this paper, we generate a commonsense knowledge graph by organizing prompts in Korean and English, respectively.

3 Evaluation

In this paper, we compute Shannon entropy and cross entropy from an information-theoretic perspective on two commonsense knowledge graphs generated with GPT-3. Table 1 shows that the indicators of the two graphs are almost the same, because the two graphs used some of the completely identical KommonGen premises sentences, resulting in only a slight difference by the language that constitutes the input prompt.

Table 1. The Result as calculated from Shannon Entropy, Cross Entropy for the information volumes of Korean Knowledge Graph generated with GPT-3. The input prompts for each graph are Korean(D_{Ko}) and English(D_{En})

Shannon Entropy	Cross Entropy
$H(D_{Ko})=5.26$	$H(D_{Ko}, D_{En})=7.31$
$H(D_{En})=5.50$	$H(D_{En}, D_{Ko})=7.16$

4 Conclusion

In this paper, we propose a method to use the KommonGen dataset as an input to the GPT-3 generation model to construct a Korean commonsense knowledge graph. Currently, additional verification of the availability of commonsense knowledge graphs generated by the proposed method is needed, but improvement work on Korean commonsense knowledge graphs will be carried out through additional data collection and correction.

5 Acknowledge

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2020-0-00368, A Neural-Symbolic Model for Knowledge Acquisition and Inference Techniques). This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425).

References

1. Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., ... & Choi, Y. (2019, July). Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, pp. 3027-3035).
2. West, P., Bhagavatula, C., Hessel, J., Hwang, J. D., Jiang, L., Bras, R. L., ... & Choi, Y. (2021). Symbolic knowledge distillation: from general language models to commonsense models. *arXiv preprint arXiv:2110.07178*.
3. Seo, J., Park, C., Moon, H., Eo, S., Kang, M., Lee, S., & Lim, H. (2021). KommonGen: A Dataset for Korean Generative Commonsense Reasoning Evaluation. In *Annual Conference on Human and Language Technology* (pp. 55-60). Human and Language Technology.
4. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.