Knowledge Graph-Augmented Korean Generative Commonsense Reasoning

Dahyun Jung [1] **Jaehyung Seo** [1] **Jaewook Lee** [1] **Chanjun Park** [1 2] **Heuiseok Lim** [1]

Abstract

Generative commonsense reasoning refers to the task of generating acceptable and logical assumptions about everyday situations based on commonsense understanding. By utilizing an existing dataset such as Korean CommonGen, language generation models can learn commonsense reasoning specific to the Korean language. However, language models often fail to consider the relationships between concepts and the deep knowledge inherent to concepts. To address these limitations, we propose a method to utilize the Korean knowledge graph data for text generation. Our experimental result shows that the proposed method can enhance the efficiency of Korean commonsense inference, thereby underlining the significance of employing supplementary data.

1. Introduction

Generative commonsense reasoning is a constrained text generation task that enables language models to learn the capacity to generate text while considering commonsense information (Lin et al., 2019). This task requires making a sentence describing a commonplace scene using a set of given concepts. For instance, given a set of concepts like toothbrush, cup, placed, toothpaste, the models should be able to output a sentence like "A toothbrush and toothpaste are placed in a cup." While humans generally possess this commonsense knowledge, it's not intrinsically present within language models, which emphasizes the importance of this task (Davis & Marcus, 2015).

Korean CommonGen (Seo et al., 2022) is a dataset that constructs on CommonGen (Lin et al., 2019), the generative commonsense reasoning dataset, in Korean. Tasks that generate text using this dataset can consider commonsense information specific to the Korean language. This

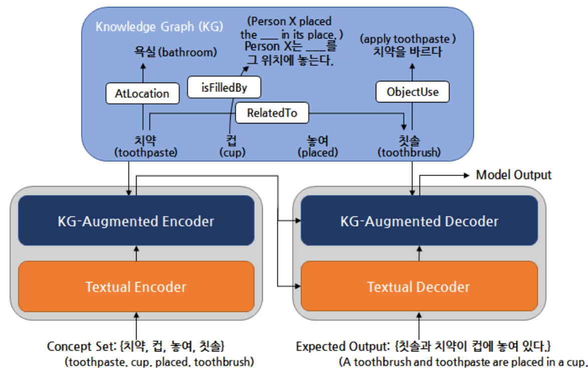

Figure 1. The illustration of knowledge graph-augmented model

dataset can be harnessed to learn commonsense knowledge that incorporates the linguistic and cultural features of the Korean language. However, the limitation of the tasks is its inability to capture the relationships between concepts absent from the dataset. The dataset does not cover every possible concept or scenario, leading to gaps in the commonsense knowledge that the model can learn. Additionally, the dataset lacks depth in certain areas, potentially leading to a superficial understanding of concepts.

In this study, we mitigate the limitation by proposing a method to augment commonsense information in generative commonsense reasoning tasks. By leveraging the Korean commonsense knowledge graph, Ko-ATOMIC, and applying it to text generation, we can take into account not only the existing commonsense relationships between concepts, but also the deeper knowledge embedded in them. We construct a model that can maximize the utilization of this knowledge graph and conduct experiments. The results intuitively demonstrate that our approach is suitable for the task of generative commonsense reasoning in Korean. This research lays the foundation for leveraging language-specific resources to enhance the capability of artificial intelligence.

2. Knowledge Graph-Augmented Commonsense Reasoning

2.1. Data Construction

Text generation data for commonsense reasoning We use the Korean CommonGen data for generative common-

[1]Department of Computer Science and Engineering, Korea University, Seoul 02841, Korea [2]Upstage, Gyeonggi-do, Korea. Correspondence to: Heuiseok Lim <limhseok@korea.ac.kr>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

Model	BLEU-3	BLEU-4	ROUGE-2	ROUGE-L	METEOR	mBERTScore	KoBERTScore
KoBART	39.54	29.16	53.6	68.55	51.17	**87.41**	92.59
mBART	**41.83**	**31.63**	**54.21**	68.36	52.08	87.25	92.26
mBART-50	40.51	30.2	53.5	68.18	50.9	87.31	92.26
KoBART (Ours)	41.13	30.23	54.02	**70.1**	**57.62**	87.33	**92.79**
mBART (Ours)	36.16	25.55	50.69	67.83	54.78	86.36	90.58
mBART-50 (Ours)	36.85	26.26	51.91	68.67	56.2	86.77	90.83

Table 1. Automatic evaluation of generation quality. The first group of models is baseline models, while the second group is our proposed knowledge graph-augmented models. The best models are **bold**, and the second best ones are underlined within each metric.

sense inference in Korean. This data considers the linguistic characteristics of the Korean language, including uniquely Korean socio-cultural terms.

Commonsense knowledge graph The most well-known existing commonsense-based knowledge graph is ATOMIC (Sap et al., 2019). In this study, we utilize the Ko-ATOMIC [1] graph, which is a translation of the existing English-based ATOMIC.

2.2. Enhancing Model with Knowledge

We aim to improve commonsense reasoning by utilizing the methodology of KG-BART (Liu et al., 2021), a knowledge graph augmented pre-trained language generation model.

The KG-BART model is differentiated from previous pre-trained language generation models by incorporating a knowledge graph, an important source of rich relational information between commonsense concepts. By utilizing this knowledge graph, the model is able to incorporate complex relationships between concepts into the text generation process. This implies that the generated sentences are not only grounded in learned linguistic patterns, but are also infused with structured knowledge, resulting in more logical text.

KG-BART consists of an encoder-decoder architecture that takes text concepts and knowledge graph as input (Figure 1). The encoder and decoder are supplemented with a KG-augmented Transformer (Vaswani et al., 2017) module based on a graph attention mechanism for incorporating entity-oriented knowledge information into the token representation. This feature allows models to capture the inherent structural correlations of intra-concept and inter-concept in the graph.

2.3. Results

We experiment with an improved methodology by applying a Korean knowledge graph. The metrics include BLEU (Papineni et al., 2002), ROUGE (Lin, 2004), METEOR (Banerjee & Lavie, 2005), and BERTScore (Zhang et al., 2019) using KoBERT [2] and mBERT (Libovickỳ et al., 2019). We compare the baseline models, KoBART [3], mBART (Liu et al., 2020), with the model enhanced by our proposed approach. mBART is a sequence-to-sequence encoder pre-trained on large corpora in multiple languages using BART (Lewis et al., 2019). KoBART is trained on a large corpus of millions of Korean sentences collected from Wikipedia, news, etc. mBART-50 is pre-trained for 50 languages using mBART's checkpoints.

In Table 1, we present the results of the automatic evaluation of generation quality for different models. The model that incorporates the knowledge graph augmentation shows a noticeable improvement in several metrics. Our improved KoBART model yields significantly better results in the ROUGE-L. It also surpasses all other models in the METEOR and KoBERTScore, achieving 57.62 and 92.79, respectively. This suggests that the model's language generation quality is highly accurate. Our proposed mBART and mBART-50 models also demonstrate a substantial increase in their performance. Even though they didn't achieve the highest scores, the improvements in their metrics are promising, highlighting the potential and effectiveness of the structured data-enhanced approach. In conclusion, our knowledge graph-augmented models exhibit notable improvements over the baseline models in several evaluation metrics, underlining the viability of our proposed method in enhancing the quality of the generated text.

3. Conclusion

In this paper, we explored the utility of the Korean commonsense knowledge graph in advancing the ability of commonsense reasoning in the Korean language. By considering relationships and in-depth knowledge, we managed to illustrate an improvement in inference quality. We demonstrated that using these graphs as an adjunct to commonsense reasoning systems can help bridge the gaps in knowledge.

[1] https://github.com/jooinjang/Ko-ATOMIC

[2] https://github.com/SKTBrain/KoBERT

[3] https://github.com/SKT-AI/KoBART

Acknowledgements

This research was supported by the Core Research Institute Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2021R1A6A1A03045425). This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2022-0-00369, (Part 4) Development of AI Technology to support Expert Decision-making that can Explain the Reasons/Grounds for Judgment Results based on Expert Knowledge).

References

Banerjee, S. and Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pp. 65–72, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics. URL `https://aclanthology.org/W05-0909`.

Davis, E. and Marcus, G. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9):92–103, 2015.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., and Zettlemoyer, L. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*, 2019.

Libovickỳ, J., Rosa, R., and Fraser, A. How language-neutral is multilingual bert? *arXiv preprint arXiv:1911.03310*, 2019.

Lin, B. Y., Zhou, W., Shen, M., Zhou, P., Bhagavatula, C., Choi, Y., and Ren, X. Commongen: A constrained text generation challenge for generative commonsense reasoning. *arXiv preprint arXiv:1911.03705*, 2019.

Lin, C.-Y. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL `https://aclanthology.org/W04-1013`.

Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., and Zettlemoyer, L. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8: 726–742, 2020.

Liu, Y., Wan, Y., He, L., Peng, H., and Philip, S. Y. Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 6418–6425, 2021.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics. doi: 10.3115/1073083.1073135. URL `https://aclanthology.org/P02-1040`.

Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N. A., and Choi, Y. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 3027–3035, 2019.

Seo, J., Lee, S., Park, C., Jang, Y., Moon, H., Eo, S., Koo, S., and Lim, H. A dog is passing over the jet? a text-generation dataset for Korean commonsense reasoning and evaluation. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pp. 2233–2249, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-naacl.172. URL `https://aclanthology.org/2022.findings-naacl.172`.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*, 2019.