A Self-Guided Approach을 활용한 한국어 텍스트 생성 쓰기 보조 기법의 향상 방법

장동현[O], 김진수[*], 이민호(교신저자)[**]
[O]경북대학교 대학원 인공지능학과, 전자공학부,
[*]경북대학교 대학원 인공지능학과, 전자공학부,
[**]경북대학교 전자공학부 인공지능대학원 인공지능학과
e-mail: dea0433@gmail.com[O], jinsoou0517@knu.ac.kr[*], mholee@gmail.com[**]

A Self-Guided Approach to Enhance Korean Text Generation in Writing Assistants

Donghyeon Jang[O], Jinsu Kim[*], Minho Lee(Corresponding Author)[**]
[O]Graduate Dept. of Artificial Intelligence, Kyungpook National University,
[*]Graduate Dept. of Artificial Intelligence, Kyungpook National University,
[**]Graduate Dept. of Artificial Intelligence, School of Electronics Engineering, Kyungpook National University

● 요 약 ●

LLM(Largescale Language Model)의 성능 향상을 위한 비용 효율적인 방법으로 ChatGPT, GPT-4와 같은 초거대 모델의 output에 대해 SLM(Small Language Model)을 finetune하는 방법이 주목받고 있다. 그러나, 이러한 접근법은 주로 범용적인 지시사항 모델을 위한 학습 방법으로 사용되며, 제한된 특정 도메인에서는 추가적인 성능 개선의 여지가 있다. 본 연구는 특정 도메인(Writing Assistant)에서의 성능 향상을 위한 새로운 방법인 Self-Guided Approach를 제안한다.

Self-Guided Approach는 (1) LLM을 활용해 시드 데이터에 대해 도메인 특화된 metric(유용성, 관련성, 정확성, 세부사항의 수준별) 점수를 매기고, (2) 점수가 매겨진 데이터와 점수가 매겨지지 않은 데이터를 모두 활용하여 supervised 방식으로 SLM을 미세 조정한다. Vicuna에서 제안된 평가 방법인, GPT-4를 활용한 자동평가 프레임워크를 사용하여 Self-Guided Approach로 학습된 SLM의 성능을 평가하였다.

평가 결과 Self-Guided Approach가 Self-instruct, alpaca와 같이, 생성된 instruction 데이터에 튜닝하는 기존의 훈련 방법에 비해 성능이 향상됨을 확인했다. 다양한 스케일의 한국어 오픈 소스 LLM (Polyglot1.3B, PolyGlot3.8B, PolyGlot5.8B)에 대해서 Self-Guided Approach를 활용한 성능 개선을 확인했다. 평가는 GPT-4를 활용한 자동 평가를 진행했으며, Korean Novel Generation 도메인의 경우, 테스트 셋에서 4.547점에서 6.286점의 성능 향상이 발생했으며, Korean scenario Genration 도메인의 경우, 테스트 셋에서 4.038점에서 5.795 점의 성능 향상이 발생했으며, 다른 유사 도메인들에서도 비슷한 점수 향상을 확인했다.

Self-Guided Approach의 활용을 통해 특정 도메인(Writing Assistant)에서의 SLM의 성능 개선 가능성을 확인했으며 이는 LLM에 비용부담을 크게 줄이면서도 제한된 도메인에서 성능을 유지하며, LLM을 활용한 응용 서비스에 있어 실질적인 도움을 제공할 수 있을 것으로 기대된다.

키워드: LLM, SLM, Vicuna, Self-Guided Approach,ChatGPT,GPT-4,Self-instruct

I. Introduction

이 연구는 Large-scale Language Model (LLM)[1,2,3]의 비용과 리소스 문제를 해결하기 위해 GPT-3, GPT-4[5] 등 잘 알려진 LLM의 출력을 기반으로 세밀하게 조정된 작은 언어 모델(Small Language Models, SLMs)을 활용하는 Self-Guided 접근 방식을 소개합니다. 이 방법은 LLMs를 특정 도메인에서 사용되는 지표를 활용하여 초기 데이터를 평가하고, 이를 토대로 SLMs을 효율적으로 학습시킵니다. 연구 결과는 이 방식이 글쓰기 지원 분야에서 SLM의 성능을 향상시키는데 도움이 될 수 있으며, LLM 기반 응용 프로그램의 비용을 줄이면서도 성능을 유지할 수 있다는 점을 보여줍니다

II. Related Works

2.1 대규모 자연어 처리 모델

최근 기계 학습의 발전으로 자연어 처리 분야에서는 대규모 언어 모델인 ChatGPT[1], LlaMA[6], Alpaca, WizardLM, Koala 등이 등장했습니다. 이러한 모델들은 인간과 유사한 대화를 이해하고 생성하는 능력을 갖추었으며, 대용량의 인터넷 텍스트 데이터를 활용하여 다양한 작업을 수행할 수 있습니다. 이러한 모델들은 지속적인 학습과 적응을 통해 정확성과 유창성을 향상시키는 능력을 갖고 있습니다.

2.2 LLM 리소스 및 시간 제약

대규모 언어 모델(GPT-3 및 GPT-4)의 구현은 리소스 및 시간 제약이 중요합니다. 훈련에 많은 자원과 시간이 필요하며, 실시간 응답을 어렵게 할 수 있습니다. 데이터 개인 정보 보호와 저장 문제도 고려되어야 합니다. 이 연구에서는 Self-Guided 접근 방식을 제안하여 작은 언어 모델(SLM)을 활용하여 훈련의 리소스 부담을 줄이고 개인 정보 보호를 하며, 효율적인 작성 보조 도구를 구현하는 것을 목표로 합니다.

2.3 Evaluation of Vicuna

Vicuna[4]와 같은 AI 모델의 성능 평가는 언어 이해, 문맥 인식, 추론과 같은 복잡한 기능을 고려해야 합니다. 그러나 기존의 평가 기준은 이러한 모델의 성능을 충분히 평가하기에는 부족할 수 있습니다. 이에 GPT-4를 활용한 새로운 평가 프레임워크가 제안되었으며, 기사 발표, 회의 기록, 보고서, 문학, 서술 문학, 드라마 등 다양한 도메인에서 테스트를 진행했습니다. 이 방법을 통해 GPT-4는 일관된 평가 점수를 제공하고 각 점수에 대한 자세한 설명을 제공할 수 있었습니다. 그러나 GPT-4와 같은 언어 모델의 특성상 입력과는 확고한 근거가 없는 정보를 생성할 수 있다는 한계를 인정해야 합니다. 이에 이 모델들에 대한 포괄적이고 표준화된 평가 시스템 개발은 여전히 진행 중인 연구 주제입니다.

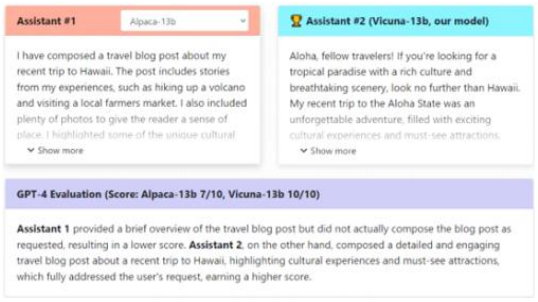

Fig. 1. Evaluation of Vicuna

III. The Proposed Scheme

한국어 작문 보조 시스템을 위해 지시 데이터를 생성하는 과정을 설명합니다. AI 허브의 다양한 데이터셋을 활용하여 다양성과 적응성을 갖춘 훈련 데이터를 구축하고, 데이터 전처리와 주석 작업을 거쳐 SLM 훈련에 적합한 형식으로 변환합니다. 이를 통해 Self-Guided 접근 방식을 통해 효과적으로 한국어 작문 보조 시스템을 훈련할 수 있습니다.

3.1 방법론

Self-Guided 접근 방식을 사용한 한국어 텍스트 생성 모델의 세부 방법론은 다음과 같은 주요 단계로 구성됩니다. 이러한 단계는 작문 보조 시스템 도메인 내에서 Small Language Model (SLM)의 성능을 향상시키는 데 중요한 역할을 합니다.

1단계: Seed 데이터 수집 및 평가

데이터 수집 단계에서는 작문 보조 시스템 도메인과 관련된 초기 훈련 데이터인 Seed 데이터를 수집합니다. 이 데이터는 Large Language Model (LLM)을 사용하여 도메인 특정 메트릭인 관련성, 유용성, 정확성 및 세부 수준에 따라 점수가 매겨집니다. 이 점수는 후속의 세밀 조정 과정에 유용한 지침 역할을 합니다.

2단계: 평가된 데이터와 평가되지 않은 데이터를 사용한 지도 학습

평가 과정 이후, 평가된 데이터와 평가되지 않은 데이터는 각각 독립적으로 SLM을 세밀 조정하는 데 사용됩니다. 평가된 데이터는 LLM에 의해 높은 관련성이 평가된 데이터로, 평가되지 않은 데이터는 보다 다양한 정보와 언어적 특성을 갖고 있습니다. 이러한 서로 다른 데이터 세트로 두 개의 모델을 훈련함으로써 성능 비교 분석이 가능해집니다.

3단계: 성능 평가

Self-Guided 접근 방식을 사용하여 SLM을 세밀 조정한 후, 그들의 성능을 평가합니다. 이 단계는 세밀 조정 과정의 효과를 평가하고 필요한 조정을 수행하기 위해 중요합니다. 성능 평가는 GPT-4를 활용한 Vicuna에서 제안된 자동 평가 프레임워크를 사용하여 수행됩니다.

LLM의 능력과 SLM의 경량성을 결합한 Self-Guided 접근 방식은 한국어 텍스트 생성 모델의 세밀 조정을 보다 집중적이고 효과적인 방식으로 가능하게 합니다. 이는 SLM의 성능을 향상시킬 뿐만 아니라

세밀 조정 과정을 보다 효율적이고 경제적으로 만듭니다.

3.2 Self-Guided 방식으로 미세 조정된 모델 평가

Self-Guided 접근 방식으로 세밀 조정된 모델을 평가하는 것은 제안된 기법의 타당성을 검증하기 위한 중요한 단계입니다. 세밀 조정 과정의 효과는 철저한 평가를 통해 확인할 수 있기 때문에, 본 섹션은 Self-Guided 접근 방식을 사용하여 훈련된 SLM의 성능을 살펴봅니다.

먼저, 모델은 Vicuna 평가 프레임워크를 통해 GPT-4 모델을 활용하여 평가되었습니다. 이 자동화된 접근 방식은 도움이 되는 정도, 관련성, 정확성 및 세부 수준이라는 평가 기준을 활용하여, 언론 보도, 회의 기록, 보고서, 문학, 서술 문학 및 드라마 등 다양한 카테고리에서 추출된 1000개의 질문에 대한 모델의 응답을 평가했습니다.

평가 결과, Self-Guided 접근 방식으로 세밀 조정된 SLM의 성능이 Self-instruct 및 Alpaca와 같은 전통적인 방법으로 훈련된 모델에 비해 크게 향상되었습니다. 이 향상은 모든 평가 메트릭에서 관찰되었으며, Writing Assistance의 맥락에서 Self-Guided 접근 방식이 SLM의 학습을 더 유용하고 관련성 있는 결과로 이끄는 데 효과적으로 작용한다는 것을 시사합니다.

하지만 GPT-4 기반 평가는 모델의 성능에 대한 유용한 통찰력을 제공하지만 한계도 있습니다. LLM이 입력에 기반하지 않는 응답을 생성하는 경향이 있어서 (hallucination이라고 함), 이는 평가 과정에서 점수에 영향을 줄 수 있습니다. 따라서 이를 해결하기 위해 평가 방법론을 더 개선해야 합니다.

또한, 이 평가는 주로 도메인 특정 작업 (Writing Assistance)에 초점을 맞추었지만, Self-Guided 접근 방식의 다른 도메인에서의 적용 가능성과 효과는 아직 탐구되어야 합니다.

평가는 또한 LLM과 비교하여 훈련에 필요한 컴퓨팅 리소스를 상당히 줄일 수 있었으며, 이는 Self-Guided 접근 방식의 비용 효율성을 강조합니다. 이는 리소스의 효율적인 활용이 가능합니다.

IV. Experiment

이 연구에서는 여섯 가지 다른 모델을 사용하여 평가를 수행했습니다. 이 중 ChatGPT와 같은 LLM 하나와 다양한 전략과 용량으로 세밀하게 조정된 다섯 가지 SLM이 포함되었습니다. 이는 점수화된 데이터를 사용하지 않고 세밀하게 조정된 13억, 38억, 58억 파라미터 모델과 Self-Guided Approach를 통해 세밀하게 조정된 38억, 58억 파라미터 모델로 구성되었습니다.

이러한 모델들은 자기 소개서, 시, 블로그, 소설, 시나리오, 뉴스라는 여섯 가지 범주에서 평가되었습니다. 각 범주에 대해 모델들은 관련성, 유용성, 정확성, 그리고 세부 수준이라는 네 가지 도메인 특정 평가 지표를 기반으로 점수를 받았습니다.

실험 결과는 Self-Guided Approach로 세밀하게 조정된 모델이 모든 범주에서 세밀하게 조정되지 않은 모델보다 일관적으로 우수한 성능을 보였음을 나타냅니다. 특히, Self-Guided Approach로 세밀하게 조정된 58억 파라미터 모델은 '자기 소개서', '블로그', '소설', 그리고 '뉴스' 범주에서 가장 높은 점수를 기록하여, 세밀하게 조정된 데이터의 통합이 이러한 특정 도메인에서 모델의 성능을 크게 향상시키는 것을 보여줍니다.

대규모 언어 모델인 ChatGPT가 가장 높은 점수를 보였습니다. 그러나 Self-Guided Approach로 세밀하게 조정된 58억 파라미터 모델이 근접하여, Self-Guided Approach의 사용으로 LLM과 SLM 사이의 성능 격차를 크게 줄일 수 있음을 보여줍니다.

전반적으로, 실험은 Self-Guided Approach 방법의 효과를 입증하고, Writing Assistance 분야에서 전통적인 세밀 조정 방법보다 우수한 성능을 나타내었습니다. 특히, 도메인 특정 점수화된 데이터를 세밀 조정 과정에 통합함으로써 모델의 성능을 다양한 범주에서 크게 향상시킬 수 있다는 것을 확인했으며, 이는 경제적이고 우수한 성능을 갖춘 쓰기 도우미 개발을 위한 유망한 전략으로 입증되었습니다.

	self	poetry	blog	novel	scenario
ChatGPT	8.755	7.966	8.552	7.942	7.967
SLM fine-tuned with 1.3 billion(no scoring data used)	3.621	3.448	4.245	4.156	3.212
SLM fine-tuned with 3.8 billion(no scoring data used)	4.284	4.24	4.921	4.623	3.967
SLM fine-tuned with 5.8 billion parameters (no scoring data used)	4.705	3.907	4.904	4.547	4.038
SLM fine-tuned with 3.8 billion parameters (scoring used)	5.815	5.997	6.129	6.036	5.68
SLM fine-tuned with 5.8 billion parameters (scoring used)	6.768	5.853	6.477	6.286	5.795

표 4.1: 13억 개(채점 데이터 제외), 38억 개(채점 데이터 포함), 58억 개(채점 데이터 제외), 38억 개(채점 데이터 포함), 38억 개(채점 데이터 포함), 5. 80억 개(채점 데이터 포함), 여기서 카테고리는 학습에 사용되지 않은 자기소개서, 시이며, 학습에 사용된 카테고리는 블로그, 소설, 시나리오, 뉴스

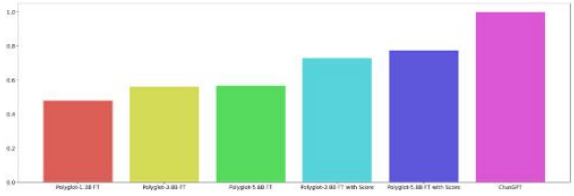

Fig. 4.1.: 모델 overview

V. Conclusions

결론적으로, 이 연구는 Writing Assistance 분야에서 Small Language Models (SLMs)의 성능을 향상시키는 새로운 방법인 Self-Guided Approach의 유효성을 성공적으로 입증했습니다. Self-Guided Approach는 Large-scale Language Models (LLMs)의 능력을 활용하여 도메인에 특화된 평가 기준을 사용하여 시드 데이터를 점수화하고, 이를 이용하여 SLMs를 지도 학습으로 세밀하

게 조정하는 데 활용합니다. GPT-4 모델을 활용한 자동 평가 프레임워크를 통해 Self-Guided Approach의 효과를 평가했습니다.

평가 결과는 Self-Guided Approach가 Self-instruct 및 alpaca와 같은 기존의 훈련 방법보다 우수한 성능을 보였습니다. 이는 Writing Assistance 분야에서 SLM의 성능을 향상시키기 위해 생성된 지시 데이터에 의존하는 방법과 비교하여 Self-Guided Approach의 유용성을 강조합니다.

또한, Self-Guided Approach는 다른 도메인에서도 일반화 가능성을 가지고 있습니다. 특정 도메인에 맞춰 평가 기준을 조정함으로써, 다른 분야에서도 유사한 성능 향상을 이끌어 낼 수 있을 것으로 기대됩니다. 그러나 이러한 가설을 검증하기 위해서는 추가적인 연구가 필요합니다.

또 다른 중요한 발견은 비용 효율성과 관련이 있습니다. 대규모 언어 모델 (LLMs)을 활용하는 서빙 및 추론 작업은 많은 비용을 소요할 수 있습니다. 그러나 Self-Guided Approach를 활용함으로써, LLM의 능력을 효과적으로 활용하면서 비용을 관리할 수 있는 잠재력이 제시됩니다. 이는 실용적인 응용 프로그램인 Writing Assistance와 같은 영역에서 자연어 생성 모델의 활용성과 확장 가능성을 높이는 데 도움이 될 것입니다.

ACKNOWLEDGEMENT

This work was partly supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1A5A7026673) and the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2021R1A2C3011169).

REFERENCES

[1] OpenAI. ChatGPT: Optimizing language models for dialogue., 2022.

[2] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. arXiv preprint arXiv:1901.02860, 2019. 1

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[4] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. Vicuna: An open-source chatbot impressing GPT-4 with 90%* chatgpt quality, 2023.

[5] OpenAI. Gpt-4 technical report. ArXiv, abs/ 2303.08774, 2023.

[6] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. Stanford Alpaca: An instruction-following LLaMA model, 2023.