LLM 을 활용한 이지리드 변환: 문장 가독성과 의미 유사성 측면에서의 문장 변환 능력 조사

정수채[1], 황재형[1], 김혜수[2], 이민지[3], 이예지[4], 이재훈[5]
[1]동국대학교 컴퓨터공학전공 학부생, [2]동국대학교 경제학과 학부생
[3]동국대학교 정보통신공학과 학부생, [4]동국대학교 영화영상학과 학부생
[5]LG U+

zmfhtmgjsej@dgu.ac.kr, ghkdwogud852@gmail.com, dddol0914@dgu.ac.kr, mingle3131@naver.com, hjyj1357@naver.com, lgjaehun@naver.com

Easy-Read Conversion Using LLM: Investigating Sentence Transformation Ability in Terms of Readability and Semantic Similarity

Su-Chae Jeong[1], Jae-Hyeong Hwang[1], Hye-Su Kim[2], Min-Ji Lee[3], Ye-Ji Lee[4], Jae-Hun Lee[5]
[1]Dept. of Computer Science and Engineering, Dong-guk University
[2] Dept. of Economics, Dong-guk University
[3]Dept. of Information and communication Engineering, Dong-guk University
[4]Dept. of Film, Dong-guk University, [5]LG U+

요 약

본 연구는 경계선 지능인 등을 대상으로 한 사기 및 부당 계약 문제를 해결하기 위해, 대규모 언어 모델(LLM)을 활용한 쉬운 말(이지리드) 변환 성능을 평가하였다. 이를 위해 Gunning Fog Index, 문장 복잡도, KoBERTScore 등의 지표로 가독성과 의미 유사도를 분석하는 평가 방법론을 제안하였으며, 여섯 종류의 LLM 을 평가한 결과 Claude-3.5-Sonnet 모델에서 우수한 성능을 확인하였다.

1. 서론

최근 경계선 지능인과 노인을 대상으로 한 사기나 부당 계약 사례가 증가하고 있다. 문서를 쉬운 언어로 변환하는 기술은 이런 문제를 해결하는 데 도움을 줄 수 있으며, 특히 대규모 언어 모델(LLM)을 활용한 쉬운 말(이지리드) 변환 기술은 문제 해결에 중요한 역할을 할 수 있다[1].

본 연구에서는 한국어 문서의 가독성을 향상시키기 위해 LLM 의 이지리드 변환 능력을 평가하고, 이를 통해 문서의 가독성 뿐만 아니라 의미 유사성을 유지할 수 있는 모델을 찾아내는 것을 목표로 한다. 이지리드 변환은 문장을 쉽게 바꾸면서도 원문의 의미를 유지해야 한다. 따라서, 이해하기 어려운 텍스트로 구성된 기계 독해 데이터를 바탕으로 LLM 이 한국어 문서를 효과적으로 변환할 수 있는지 검토하고, 이지리드 변환 검증을 위한 평가 방법론을 제시하고자 한다.

또한 평가를 위해 사용된 코드와 데이터셋을 공개하여 한국어 이지리드 변환에 대한 접근과 연구를 촉진하고자 한다.[1]

[1] https://github.com/SChaeck/LLM-korean-readability-assessment

2. 관련 연구

최근 국어 문서의 가독성을 평가하기 위해 Gunning Fog Index, 문장 복잡도(SC) 등 다양한 지표를 사용한 연구들이 많이 이루어져 왔다[2, 3, 4]. 그러나 이러한 연구들은 주로 기존 문서를 평가하는 데 중점을 두었고, LLM 을 활용한 한국어 문서 변환 능력을 측정하는 연구는 상대적으로 부족했다.

본 연구는 한국어 문서를 대상으로 한 이지리드 변환 성능 평가를 위해, LLM 모델의 문장 변환 능력을 체계적으로 평가하고, 가독성 및 의미 유사도를 고려한 새로운 평가 방법론을 제시하고자 한다.

3. 실험 구성

다양한 분야에서 LLM 모델의 이지리드 변환 능력을 평가하기 위해, AI-Hub 에서 제공하는 한국어 문서 독해 데이터[2][3][4][5][6] 81,805 개를 '금융/법률' 100 개, '기술과학' 50 개, '뉴스 기사' 50 개, '행정 문서' 50 개, '도서자료' 50 개씩 랜덤 샘플링하여 300 개의 통합 데이터셋을 구축하였다.

[2] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=71610
[3] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=71533
[4] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=577
[5] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=569
[6] https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihubDataSe=data&dataSetSn=92

실험에는 '한국어 언어모델 다분야 사고력 벤치마크'인 LogicKor [7] 에서 상위권을 차지한 'GPT-4o', 'Gemini-Pro-1.5', 'Claude-3-5-sonnet'을 사용하였다. 또한, 'GPT-4o'의 증류 모델인 'GPT-4o-mini', 그리고 오픈소스 sLLM 중 우수한 성능을 보인 'Ko-gemma-2-9b-it, 'Glm-4-9b-chat' 모델 또한 평가에 포함되었다.

성능 측정을 위해 '문장 복잡도(SC)'와 'Gunning Fog Index(GFI)'[4]를 성능 지표로 사용해 가독성을 평가하였고, BERTScore[5]를 한국어에 대해 적용한 KoBERTScore(KBS)를 사용해 의미 유사성을 측정했다. 또한 인간 평가와 높은 유사도를 갖는 것으로 알려진 GPT-4o 모델[6]을 사용하여 가독성 스코어(RS)를 추가 평가하였다.

<표 1> LLM의 이지리드 변환 성능 확인

Method	SC(↓)	GFI(↓)	RS(↑)	KBS(↑)
Original Paragraph	8.22	23.39	5.96	1
GPT-4o$_{0shot}$	6.95	18.22	6.85	0.68
GPT-4o$_{1shot}$	6.35	17.48	6.86	0.67
GPT-4o-mini$_{0shot}$	5.95	15.69	7.37	0.64
GPT-4o-mini$_{1shot}$	5.9	15.51	7.02	0.69
Gemini-Pro-1.5$_{0shot}$	6.68	14.71	8.38	0.46
Gemini-Pro-1.5$_{1shot}$	6.13	14.88	8.19	0.50
Claude-3.5-Sonnet$_{0shot}$	5.82	14.59	8.03	0.50
Claude-3.5-Sonnet$_{1shot}$	5.39	13.70	8.00	0.52
Ko-Gemma-2-9b-it$_{0shot}$	5.57	14.54	7.76	0.52
Ko-Gemma-2-9b-it$_{1shot}$	5.52	14.64	7.91	0.52
Glm-4-9b-chat$_{0shot}$	5.89	14.66	7.46	0.57
Glm-4-9b-chat$_{1shot}$	5.71	14.97	7.47	0.57

4. 결과 분석

지표 분석. Claude-3.5-Sonnet$_{1shot}$ 모델은 SC 와 GFI 에서 다른 LLM 에 비해 뛰어난 성능을 보였으며, 문장을 단순화하고 가독성을 향상시키는 데 우수한 능력을 보였다. 반면, GPT-4o$_{0-shot}$ 모델은 SC 와, GFI 에서 가장 낮은 성능을 보였다.

RS 에서는 Gemini-Pro-1.5$_{0-shot}$ 모델이 가장 높은 점수를 기록했으나, SC 와 GFI 에서는 상대적으로 낮은 성능을 보였다. 이는 인간 평가와 SC, GFI 기준 간에 불일치가 있을 가능성을 시사한다.

흥미롭게도 GPT-4o 계열 모델은 KBS 에서 높은 점수를 기록했는데, 이는 의미 유사도와 문장 단순화 사이의 상충 관계를 보여준다. 또한 GPT 계열 모델이 이지리드 변환에서 기존 문서의 단어를 쉽게 바꾸지 않는 경향이 있음을 확인할 수 있다.

sLLM. Ko-Gemma-2-9b-it 은 Gemeni, Claude 와 같은 LLM 과 비교하더라도 전반적으로 뛰어난 성능을 보였으며, 특히 SC, GFI 에서 높은 성적을 기록했다. 또한 GPT-4o-mini 는 KBS 에서 가장 높은 성적을 기록하면서 SC 와 GFI 에서 괜찮은 성능을 보였다. 따라서 이지리드 변환에서 sLLM 을 사용하는 것이 자원 율적일 수 있다.

0-shot, 1-shot. SC 와 GFI 지표에서는 1-shot 설정이 대부분의 모델의 성능을 향상시키는 경향을 보였다. 그러나 RS 에서는 일부 모델에서 큰 차이가 없거나, 오히려 성능이 소폭 감소하는 경향을 확인했다. KBS 결과 또한 모델마다 차이가 있어, 1-shot 설정이 이지리드 성능을 일관되게 향상시킨다는 근거를 찾을 수 없었다.

5. 결론 및 한계

본 연구에서는 여섯 종류의 LLM 에 두 가지 설정을 사용하여 한국어 문서의 이지리드 변환 능력을 평가하였다. 실험 결과, Claude-3.5-Sonnet 모델이 문장 복잡도와 Gunnig Fog Index 에서 가장 우수한 성능을 보였으며, GPT 계열 모델은 의미 유사도에서 상대적으로 높은 성적을 기록했다. 또한 문장 단순화와 의미 유사도 간의 Trade-Off 를 간접적으로 확인할 수 있었으며, sLLM 의 성능이 LLM 과 비교해도 떨어지지 않음을 확인할 수 있었다.

이 연구를 통해 한국어 이지리드 변환에 대한 여러 LLM 의 성능을 확인하였지만 한정된 데이터셋을 사용했다는 점과, 실제 인간 평가를 측정하지 않았다는 한계가 존재한다. 후속 연구에서는 다양한 문서 유형에 대한 추가 실험과 인간 평가 및 인간 평가와의 일치성을 높이는 측정 지표 도입이 필요할 것으로 생각된다.

Acknowledgement

※ 본 논문은 과학기술정보통신부 대학디지털교육 역량강화 사업의 지원을 통해 수행한 ICT 멘토링 프로젝트 결과물입니다.

※ 본 논문은 현대차 정몽구 재단 장학생으로서 지원을 받아 수행된 연구입니다.

참고문헌

[1] Freyer, N., Kempt, H. & Klöser, L., "Easy-read and large language models: on the ethical dimensions of LLM-based text simplification," Ethics Inf Technol, vol. 26, no. 3, pp. 50, 2024.

[2] 조찬우, 조찬형, 우균, "가독성 평가를 위한 한글 문장의 복잡도 측정 방법," 한국정보과학회 2018 한국컴퓨터종합학술대회, 제주, 2018, pp. 2265-2267.

[3] 정대영, "소설 텍스트의 문장 복잡도 연구-자동화된 프로그램을 활용하여-," 문학교육학, no. 48, pp. 263-292, 2015.

[4] 김미란, "대학수학능력시험 영어 독해지문의 어휘 다양성 및 가독성 분석," 외국어교육연구, vol. 36, no. 4, pp. 71-90, 2022.

[5] Zhang Tianyi, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, Yoav Artzi, "BERTScore: Evaluating Text Generation with BERT.," In Proceedings of the 8th International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia, 2020

[6] Sottana, Andrea, Bin Liang, Kai Zou, and Zheng Yuan, "Evaluation Metrics in the Era of GPT-4: Reliably Evaluating Large Language Models on Sequence to Sequence Tasks," In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (ICLR), Singapore, 2023, pp. 8776–8788.

[7] https://lk.instruct.kr/