

2025

스마트미디어 추계학술대회

2025 Autumn Academic Conference of Smart Media

AI 대전환: 인간, 미디어, 기술의 새로운 공존 패러다임

AI Transformation: A New Coexistence Paradigm of
Human, Media and Technology

일시 | 2025년 11월 14일(금) ~ 16일(일)

장소 | 국립 군산대학교 (황룡문화관)

구두 발표 O-13

● 시 간	2025년 11월 15일 (토) 15:10~16:20
● 발표장	황룡문화관 3층 306호
● 좌 장	김영철 교수 (홍익대)

013-01	제 목 Two remedies for classifying Alzheimer's disease within EfficientNet B0 pipeline 저 자 수바시 루이텔, 정성원, 김지인, 변재영, 권구락 (조선대)	page 107
013-02	제 목 문법 증강을 통한 인공지능의 자연어 이해 성능 향상 저 자 양진모, 진예진, 이상호, 김영철 (홍익대)	page 109
013-03	제 목 Genome-Wide Association and Genomic Prediction of Heat Stress Tolerance in Olive Flounder (<i>Paralichthys olivaceus</i>) 저 자 H.A.C.R. 한차풀라, D.S. 리예나게, W.K.M. 오메카, 김가은, 김정은, 이지훈, Y.K. 코다고다, M.A.H. 딜샨, D.C.G. 로드리고, G.A.N.P. 가네풀라, 아석경, 이제희 (제주대)	page 111
013-04	제 목 Compilation of primary disease-related data in olive flounder for diagnostic app design 저 자 Cheonguk Park, Hanchang Sohn, Changnam Jin, Taehyug Jeong, Jehee Lee (제주대)	page 114

문법 증강을 통한 인공지능의 자연어 이해 성능 향상

양진모¹, 진예진², 이상호³, 김영철^{4*}
 홍익대학교 소프트웨어공학연구실
 e-mail : {¹ljmd2222, ²yejin_jin}@g.hongik.ac.kr,
³project@rastech.co.kr, ^{4*}bob@hongik.ac.kr

Enhancing performance of natural language understanding of AI with syntax augmentation

Jinmo Yang¹, Yejin Jin², Sangho Lee³, R. Young Chul Kim^{4*}
 SE Lab., Hongik University

요약

현재 인공지능을 지속적으로 발전시키며, 잠재된 능력을 끌어내기 위한 연구가 활발히 되고 있다. 그중 대규모 언어 모델(LLM)에 대해 프롬프트 기반 질의(Query)로 더 좋은 결과를 얻고자 한다. 하지만 기초 단위에서의 간결성과 명확성, 즉 질의 속 문장의 종류(복문/중문/단문) 측면에서 이를 연구한 사례는 매우 적다. 이를 위해, 자동 문장 구조 변형 기반 텍스트 처리 메카니즘을 제안한다. 이는 문어체/구어체의 단문, 중문, 그리고 복문으로 확장하여, 다양한 질의를 LLM의 입력과 비교하여 좋은 질의를 제시하고자 한다. 비교 성능 평가 결과를 볼 때, 내림차순으로 단문, 원문, 중문, 구어체, 그리고 복문으로 나타난다. 이 결과로 프롬프트를 구성할 때의 문장 구성에 대한 지표 자료 역할을 기대한다.

1. 서론

현재도 자연어는 단어나 표현의 선택이 자유롭고 맥락에 강하게 의존하기 때문에 모호하다. 따라서 인공지능이 이를 이해하기 위해 형태소 및 구문 분석, 의미 분석, 맥락과 시대에 따른 유행어 학습 등의 방법이 있다[1]. 또한 대규모 언어 모델(LLM)에 사용되는 어텐션은 자연어 입력이 들어왔을 때 집중해야 할 부분에 비중으로 의미를 더 잘 이해시키고자 한다[1]. 또한 사용자의 의도를 더욱 잘 파악하기 위해 프롬프트 기법들을 사용하고, 학습 외의 도움을 주기 위해 검색증강생성(RAG) 방법으로 자연어 생성에 필요한 데이터를 조회하여 질이 높은 결과를 얻고자 한다[2]. 이렇게 다양한 방법으로 인공지능의 성능을 높이는 노력이 있지만, 그 중 문법 증강을 통해 성능을 높이는 시도는 매우 적다. 특히 동일한 맥락에서 다양한 문법에 따른 성능을 확인하는 연구는 희박하다. 따라서 본 논문은 주어진 자연어 입력을 문어체의 단문, 중문, 복문, 그리고 구어체로 증강했을 때의 LLM의 성능을 확인한다. 2장에서는 관련 연구, 3장에서는 자연어 증강 프로세스, 4장에서는 모델 학습 및 성능, 그리고 5장에서는 결론을 언급한다.

2. 관련 연구

2.1 미완성/단어 순서 바꿈을 통한 한국어 능력 평가

문법과 관련된 인공지능의 자연어 이해에 대한 연구로, Kim 등[3]은 한국어 이해 능력 평가(KLUE)[4] 벤치마크 데이터셋으로부터 문법적으로 불완전한 한국어(SIKO) 데이터셋을 생성했다. 이는 문장 부호 삭제, 의미를 유지/삭제되도록 단어의 순서 바꾸기, 그리고 이것들의 조합으로 구성되어 있다. 이 데이터셋으로 기존 LLM을 평가 및 학습 후 평가를 진행했을 때 불완전성의 종류와 정도에 따라 성능이 다르게 평가되는 것을 확인했

다. 문법을 불완전하게 만드는 것에 대비해 우리는 질의 문장 구조 기반 다른 데이터, 즉 단문, 중문, 복문, 그리고 구어체의 자연어 입력에 따른 LLM의 자연어 이해 능력을 평가하고자 한다.

2.2 자동 자연어 문법 증강

본 논문의 선행 연구로 자동 자연어 문법 증강이 있다[5]. 이는 LLM을 통한 자동 문장 증강 프로세스로, 프롬프트 기법들을 통해 LLM을 한국어 문법에 특화되도록 하고, 특화된 LLM을 반복문 속에서 입력되는 자연어를 단문, 중문, 복문, 그리고 구어체로 증강하도록 한다. 생성된 문장을 검증하는 과정이 필요한데, LLM을 통해 검증을 진행하려 한다.

2.3 3D장면 저작 도구의 장면 도메인 문장 생성

또한 3D장면 저작 도구 제작 목표[6]를 위해 저자들은 만화에 알맞은 도메인에서의 서술 문장들을 생성하고 이에 대한 인공지능의 이해도를 평가하는 중이다[7]. 이에 따라 수작업으로 문장을 생성하는 것과 인공지능을 활용한 문장의 생성이 있다. 수작업으로 생성한 문장 558개에 대해서는 머신러닝 모델 랜덤포레스트 분류기의 학습 결과 0.984의 f1-score를 기록했다. 프롬프트 기법들과 구조화된 출력(structured output) 기능을 통해 10개 도메인(사무실, 학교, 카페, 슈퍼, 동물원, 미용실, 대기실, 매점, 지하철, 놀이공원)에서의 문장을 생성 중이다.

3. 자연어 증강 프로세스 개선

원본 한국어 데이터는 KLUE 벤치마크 데이터셋의 토픽 분류(topic classification) 데이터로, 한국어 뉴스 기사의 제목으로부터 뉴스의 도메인(정치, 경제, 사회, 문화, 세계, IT/과학, 스포츠)를 예측하는 데이터셋이다[4]. 이 원본 데이터를 한 번의 프롬프트 엔지니어링 과정을 통해 단문, 중문, 복문, 그리고 구어체의

문장으로 증강한다.

이 데이터가 올바로 증강되었는지 확인하고 개선되어야 한다면 개선하기 위해 검증 과정을 거친다. 이를 위해 구문분석에 대한 프롬프트 엔지니어링을 거친 LLM에 증강된 문장들을 입력하여 실제 문장 형태에 알맞게 수정한다. 그림 1은 전체 자연어 증강 프로세스 개선을 표현한다.



(그림 1) KLUE 토픽 분류 데이터 증강 프로세스 개선

H200 141GB GPU 서버에서 Qwen 3 32B[8] 모델을 프롬프트 엔지니어링 후 45,678개 뉴스 제목을 증강하여 원본 뉴스 제목 포함 총 228,390개의 자연어 데이터포인트를 수집하였다. 다음 장에서 이를 학습시키고 분석한 내용을 다룬다.

4. 증강된 데이터 학습 및 평가

증강된 데이터로 인공지능을 학습시키는 내용을 다룬다. 학습 서버 스펙은 H200 141GB GPU이고, 모델 구성은 Qwen 3 32B[8] 모델에 대해 QLoRA로 양자화 및 LoRA 어댑터를 사용한다. 데이터의 비율은 각 데이터셋의 45,678개의 데이터에 대해 학습 데이터는 75 %과 검증 및 평가 데이터는 각각 12.5 %로 분할한다.

(표 1) 문법 증강 자연어 학습 모델 성능

학습한 데이터	평가	성능 (f1-score)
제목	제목	0.8164689689817143
	단문	0.8197980730795493
	중문	0.8136007326269146
	복문	0.8097811018048188
	구어체	0.8187144187486863
단문	제목	0.8378772723780498
	단문	0.8186361174255511
	중문	0.8245266390717187
	복문	0.8154197990668963
	구어체	0.8165283778088762
중문	제목	0.8387863710855245
	단문	0.8320146873853379
	중문	0.7237498112947982
	복문	0.640822177224908
	구어체	0.640951085102317
복문	제목	0.5990206945019338
	단문	0.5352327258436513
	중문	0.5383390882388818
	복문	0.40875217146602894
	구어체	0.483305635153031
구어체	제목	0.7046248586073007
	단문	0.700897449497887
	중문	0.7036598931567847
	복문	0.69680434225154
	구어체	0.6951406652804828

표 1은 모델의 학습 데이터를 고정하고, 모든 종류의 데이터에 대해 평가를 진행한 바를 나타낸다. 굵게 표시한 것은 학습 데이터와 평가 데이터를 같은 종류로 한 경우이다. 이에 따르면 평가 성능은 내림차순으로 단문, 제목, 중문, 구어체, 그리고 복문으로 나타난다. 문어체인 단문, 중문, 그리고 복문을 고려할 때, 인공지능은 단순한 문장들로 구성한 입력을 더 잘 이해한다고 본다. 원문인 뉴스 기사 제목은 문법 증강 전 본래 의미를 명확화

게 담고 있어서 성능이 높은 것으로 이해할 수 있다.

반면, 학습 데이터를 고정하고 모든 종류의 평가 데이터에 대해 성능의 비슷하고 다름에 대한 것은 추가 분석이 필요하다. 즉, 제목, 단문, 그리고 구어체는 모든 종류의 평가 데이터에 대해 성능이 비슷하지만, 중문과 복문은 서로 다르다. 구문 분석 외에 별도의 데이터 품질 검증 방법을 고안할 필요가 있다.

4. 결 론

본 논문은 LLM에 자연어를 입력할 때 그 구성이 문법적으로 달을 경우 출력 성능에 대한 평가를 진행했다. 우리는 문법을 명확히 하고 문장들을 간결하게 구성했을 때, 즉 단문, 중문, 그리고 복문 순으로 입력을 더 잘 이해하여 출력을 생성하는 것을 확인할 수 있다. 이는 인공지능이 사용자의 요구사항을 더 잘 이해할 수 있도록 정제된 입력을 사용하는 것과 일맥상통한다. 향후 더 면밀하게 데이터를 검증하고, 생각/이성 생성(reasoning mode)에 대해서도 각각의 문법에 따른 성능을 확인하고, 기존 3D저작 장면의 문장과 데이터를 병합할 예정이다.

감사의 글

본 연구는 2024년도 문화체육관광부의 지원으로 한국콘텐츠진흥원(과제명: 인공지능 기반 사용자 대화형 멀티모달 인터랙티브 스토리텔링 3D장면 저작 기술 개발, 과제번호: RS-2023-00227917, 기여율: 50 %) 지원과 한국연구재단의 4단계 두뇌한국21사업(과제명: 초분산 자율 컴퓨팅 서비스 기술 연구팀, 과제번호: 202003520006, 기여율: 50 %)의 지원을 받아 수행된 연구임.

참고 문헌

- [1] R. Patil, et al., "A survey of text representation and embedding techniques in nlp," IEEE Access, Vol.11, Apr. 2023, pp.36120–36146.
- [2] Louie Giray, "Prompt engineering with ChatGPT: a guide for academic writers," Annals of biomedical engineering, Vol.51, Jun. 2023, pp.2629–2633.
- [3] J. Kim, et al., "Does incomplete syntax influence Korean language model? Focusing on word order and case markers," arXiv, Jul. 2024, pp.1–19.
- [4] S. Park, et al., "KLUE: Korean language understanding evaluation," arXiv, Nov. 2021, pp.1–76.
- [5] J. Yang, et al., "Automatic textual data transformation for enhancing f1-score on classification," ISAAC 2024, Vol.24, Nov. 2024, pp.2–6.
- [6] 문화체육관광부. "인공지능 기반 사용자 대화형 멀티모달 인터랙티브 3D 장면 저작 기술 개발." 문화체육관광 연구개발사업 신규 기획과제, 2022.
- [7] J. Yang, et al., "Topic classification training model with automatic textual data transformation," ICGHIT 2025, Jan. 2025, pp.78–81.
- [8] Qwen Team, "Qwen3 technical report," arXiv, May 2025, pp.1–35.