Wilson Score 기반 상위 댓글 선별과 DPO 선호 최적화를 활용한 한국어 뉴스 댓글 생성 모델

길기훈, 홍수민, 박천음, 이상금*
*국립한밭대학교

minegihun@gmail.com, ghdtnaos@gmail.com, parkce@hanbat.ac.kr, *sangkeum@hanbat.ac.kr

A Model for Selecting Top Comments Based on Wilson Score and Generating Korean News Comments Using DPO Preference Optimization

Gihun Gil, Sumin Hong, Cheoneum Park, Sangkeum Lee*
*Hanbat National Univ.

요 약

최근 생성형 AI 의 기술적 고도화로 기계가 생성한 인간이 작성한 텍스트의 구분이 모호해짐에 따라, 허위 정보 유포 및 여론 조작의 여지가 있다. 본 논문에서는 한국어 특화 거대 언어 모델(LLM)을 기반으로 뉴스 기사의 문맥을 심층적으로 이해하고 자연스러운 댓글을 생성하는 시스템을 제안한다. 20 만 건의 뉴스 데이터를 활용하여 신뢰구간 95%의 Wilson Score 을 적용하여 SFT(Supervised Fine-Tuning)를 수행하고, 감정 분석 모델을 활용한 DPO(Direct Preference Optimization)를 적용하여 선호도 기반 윤리적 정렬을 수행한다. 실험 결과, 제안된 모델은 인간 데이터 대비 낮은 PPL(Perplexity)수치를 보였으며, 의도된 감정 분포를 따르는 댓글을 생성함을 확인한다.

Ⅰ. 서 론

생성형 AI 의 급속한 발전으로 온라인 뉴스와 소셜 미디어에서 AI 생성한 허위 정보와 댓글의 생산·유통이 증가하고 있다. 한 연구에 따르면 일반인이 AI 생성 댓글의 67%를 사람이 작성한 것으로 오인할 정도로 비전문가가 진위를 판별하기 어려운 수준에 이른다[1].

본 논문에서는 AI 생성 댓글을 식별할 수 있는 판별 모델을 학습시키기 위해 한국어 능력이 우수한 LLM 을 기반으로 뉴스 기사의 문맥을 심층적으로 이해하고 실제 인간의 댓글 패턴을 정교하게 모사하는 뉴스 댓글 생성 모델을 설계하여 인간 데이터의 대조군으로 제공한다[2].

Ⅱ. 본론

2.1. 베이스 모델 선정

뉴스 기사의 문맥을 정확히 파악하고, 알맞은 한국어 댓글을 생성하기 위해 한국어 텍스트에 대한 이해도가 높은 사전학습 모델을 선정한다. 본 연구에서는 Kanana-2.1B, Kanana-8B, 그리고 KorMo-10B 를 베이스 모델로 채택한다[3][4].

2.2. 데이터 수집 및 데이터셋 구축

2024 년 4 월부터 2025 년 9 월까지 네이버 뉴스 일별 최다 댓글 기사를 중심으로 약 20 만 건의 뉴스 데이터를 수집한다.

Supervised Fine-Tuning(SFT) 데이터셋 구축을 위해 댓글의 신뢰구간 95%의 Wilson Score 의 하한값을 적용하여 표본 수가 적을 때 발생할 수 있는 인기도 편향을 보정하여 뉴스별로 가장 신뢰도 높은 상위 댓글을 선별하여 총 55,262 건의 데이터를 확보한다[5]. 표본 n 개에 대해 유의수준 a 에서의 Wilson Score 의 하한값은 다음과 같은 식으로 계산한다.

$$p = \frac{\hat{p} + \frac{z_{a/2}^2}{2n} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{a/2}^2}{4n^2}}}{1 + \frac{z_{a/2}^2}{n}}$$

2.3. DPO 를 통한 선호도 기반 윤리적 정렬

Direct Preference Optimization(DPO)은 NSMC(네이버 영화 리뷰 데이터)로 파인튜닝된 BERT 모델을 이용해

각 뉴스 댓글의 감정을 분석하고 부정 확률이 가장 낮은 댓글을 선호(Chosen), 가장 높은 댓글을 비선호(Rejected)로 설정하여 22,788 개의 데이터를 확보한다.

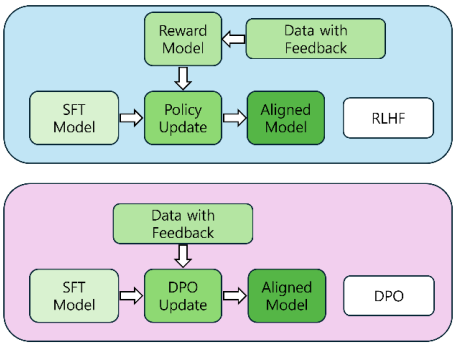

그림 1. RLHF 와 DPO 의 학습 파이프라인 비교

모델이 앞서 구축된 데이터셋의 의도(정제된 표현)에 부합하는 댓글을 생성하도록 DPO 를 수행한다. 그림 1 과 같이 기존의 Reinforcement Learning from Human Feedback(RLHF) 방식과 달리 DPO 는 별도의 Reward Model 없이 언어 모델 자체를 Reference Model 로 사용하여 선호 데이터(Chosen)와 비선호 데이터(Rejected)의 선호도를 조정하는 방식으로 최적화를 수행한다[6].

2.4. 실험 환경 및 결과 분석

	SFT	DPO
Batch size	**16**	
Learning rate	**2e-4**	**5e-6**
Epochs	**3**	
Sequence length	**4096**	
DPO beta	**-**	**0.1**

표 1. SFT & DPO 학습 파라미터

	PPL(DPO)	NEG_P(SFT)	NEG_P(DPO)
Human	**1027.69**	**0.36(chosen) / 0.87(rejected)**	
Kanana-2.1B	**22.65**	**0.63**	**0.02**
Kanana-8B	**18.19**	**0.63**	**0.03**
KORMo-10B	**32.58**	**0.63**	**0.14**

표 2. PPL 와 성능 평가 지표

모델 학습 데이터로서의 유용성을 판단하기 위해 표 1 의 파라미터를 사용하여 모델을 학습한 후, 생성된 댓글에 대해 Perplexity(PPL)과 부정 확률을 분석하여 표 2 에 정리한다. PPL 의 경우 AI 생성 결과가 Human 데이터 대비 낮은 점수를 나타내며, AI 생성 텍스트가 통계적으로 매우 전형적인 패턴을 따른다. 또한 제안된 모델들은 DPO 학습을 통해 선호도 기반 윤리적 정렬이 완료한다. SFT 학습 후의 부정 확률(NEG_P)은 세 모델 전부 약 0.63 이고, DPO 적용 후에는 세 모델 모두 Chosen 데이터보다 낮은 수치를 나타낸다. KorMo-10B 의 경우 Kanana 기반 모델과 달리 상대적으로 높은 부정확률을 나타내고, 이는 높은 파라미터로 인해 모델이 단순히 긍정 어휘만을 반복하도록 과적합되지 않았기 때문으로 해석된다. 결과적으로, 다양한 뉴스 맥락을 반영해야 하는 판별 모델의 학습용 데이터 생성 측면에서는 KorMo-10B 가 더 유리한 특성을 보인다고 판단할 수 있다.

Ⅲ. 결론

본 논문에서는 고성능 AI 생성 텍스트 판별 모델의 학습 데이터 확보를 위해, 뉴스 기사의 문맥을 이해하고 알맞은 댓글을 생성하는 뉴스 댓글 생성 모델을 제안한다. 신뢰도 높은 데이터 학습을 위해 Wilson Score 기반의 SFT 와 감정 정보 기반의 DPO 를 단계적으로 적용하였으며, 실험 결과 제안된 모델들은 통계적으로 안정적인 텍스트를 생성하며, DPO 적용 후 부정 확률이 유의미하게 감소하여 모델의 선호도 기반 윤리적 정렬이 효과적으로 이루어졌음을 입증한다. 향후 연구에서는 구축된 고품질의 AI 생성 댓글 데이터셋과 실제 인간 댓글 데이터를 활용하여, 최신 LLM 이 생성한 텍스트까지 정밀하게 탐지할 수 있는 딥러닝 기반 판별 모델을 개발하고 그 성능을 검증할 계획이다.

참 고 문 헌

[1] GO, Wooyoung, et al. XDAC: XAI-driven detection and attribution of LLM-generated news comments in Korean. In: *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2025. p. 22728-22750.

[2] Lee S, Nengroo SH, Jin H, Doh Y, Lee C, Heo T, Har D. Anomaly detection of smart metering system for power management with battery storage system/electric vehicle. ETRI Journal. 2023 Aug;45(4):650-65.

[3] BAK, Yunju, et al. Kanana: Compute-efficient bilingual language models. *arXiv preprint arXiv:2502.18934*, 2025.

[4] KIM, Minjun, et al. KORMo: Korean Open Reasoning Model for Everyone. *arXiv preprint arXiv:2510.09426*, 2025.

[5] WILSON, Edwin B. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 1927, 22.158: 209-212.

[6] RAFAILOV, Rafael, et al. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 2023, 36: 53728-53741.