

# 초거대 언어 모델로부터의 추론 데이터셋을 활용한 감정 분류 성능 향상

박눈솔<sup>○</sup>, 이민호(교신저자)<sup>\*</sup>

<sup>○</sup>경북대학교 데이터사이언스대학원,

<sup>\*</sup>경북대학교 대학원 인공지능학과, 전자공학부

e-mail: nunsolp@knu.ac.kr<sup>○</sup>, mhlee@gmail.com<sup>\*</sup>

## Empowering Emotion Classification Performance Through Reasoning Dataset From Large-scale Language Model

NunSol Park<sup>○</sup>, MinHo Lee(Corresponding Author)<sup>\*</sup>

<sup>○</sup>Graduate School of Data Science, Kyungpook National University,

<sup>\*</sup>Graduate Dept. of Artificial Intelligence, School of Electronics Engineering, Kyungpook National University

### ● 요약 ●

본 논문에서는 감정 분류 성능 향상을 위한 초거대 언어모델로부터의 추론 데이터셋 활용 방안을 제안한다. 이 방안은 Google Research의 'Chain of Thought'에서 영감을 받아 이를 적용하였으며, 추론 데이터는 ChatGPT와 같은 초거대 언어 모델로 생성하였다. 본 논문의 목표는 머신러닝 모델이 추론 데이터를 이해하고 적용하는 능력을 활용하여, 감정 분류 작업의 성능을 향상시키는 것이다. 초거대 언어 모델(ChatGPT)로부터 추출한 추론 데이터셋을 활용하여 감정 분류 모델을 훈련하였으며, 이 모델은 감정 분류 작업에서 향상된 성능을 보였다. 이를 통해 추론 데이터셋이 감정 분류에 있어서 큰 가치를 가질 수 있음을 증명하였다. 또한, 이 연구는 기존에 감정 분류 작업에 사용되던 데이터셋만을 활용한 모델과 비교하였을 때, 추론 데이터를 활용한 모델이 더 높은 성능을 보였음을 증명한다. 이 연구를 통해, 적은 비용으로 초거대 언어모델로부터 생성된 추론 데이터셋의 활용 가능성을 보여주고, 감정 분류 작업 성능을 향상시키는 새로운 방법을 제시한다. 제시한 방안은 감정 분류뿐만 아니라 다른 자연어처리 분야에서도 활용될 수 있으며, 더욱 정교한 자연어 이해와 처리가 가능함을 시사한다.

**키워드:** 감정 분류(Emotion Classification), 추론 데이터셋(Reasoning Dataset), 초거대 언어 모델(Large-scale Language Model), 생각 사슬(Chain of Thought)

## I. Introduction

감정 분류는 자연어 처리 분야에서 중요한 과제로, 텍스트에서 사람의 감정 상태를 인식하고 분류하는 것을 목표로 한다. 이는 고객의 리뷰 분석, 소셜 미디어 감정 분석, 감정 챗봇 등 다양한 분야에서 활용되며, 이를 위해 다양한 머신러닝 기법과 데이터셋이 사용되고 있다. 그러나 감정 분류는 여전히 도전적인 작업으로 남아있는데, 그 이유는 텍스트에서의 감정 표현이 복잡하고 다양하며 맥락적인 정보가 포함되기 때문이다. 또한 감정 분류 작업에는 데이터 레이블이 다양한 측면에서 중요한지만 레이블된 데이터셋을 생성하는 것은 많은 비용이 든다는 문제점이 있다.

최근에는 초거대 언어 모델 ChatGPT, GPT-4와 같은 모델이 많은 주목을 받고 있다. 이러한 모델은 다양한 자연어 처리 과제에서 뛰어난 성능을 보여주고 있으며, 특히 사람처럼 추론하는 능력을 가지고 있다는 점에서 주목받았다[1].

우리는 ChatGPT로부터 추론 데이터를 생성하고 이를 활용하여 감정 분류 모델을 훈련시켰다. 본 연구를 통해 비교적 적은 비용으로 초거대 언어 모델로부터 생성된 추론 데이터가 감정 분류 작업에서 어떻게 활용될 수 있는지를 탐색하고, 모델의 감정 분류 성능을 향상시키는 방법을 제안하고자 한다.

## II. Related Works

최근의 연구에서는 BERT, GPT 계열의 사전 학습된 언어 모델이 감정 분류 작업에서 뛰어난 성능을 보이고 있다[2][3][4][5]. 별개로, 추론 데이터(Reasoning Data)에 대한 연구도 진행되고 있다. Google Research의 “Chain-of-Thought Prompting Elicits Reasoning in Large Language Models”에서는 추론 데이터를 활용하여 모델의 해석력을 향상시키고자 하였다[6]. 생각 사슬(Chain-of-Thought)의 개념을 적용하여 모델의 추론 능력을 향상시켜 더 이상 모델의 크기를 늘리지 않고도 여러 복잡한 작업들을 높은 성능으로 수행하였다. 또한 ChatGPT, GPT-4와 같은 뛰어난 추론 능력을 가진 초거대 언어 모델로부터 생성된 데이터가 어떻게 활용될 수 있는지 탐색하는 것도 최신 연구 중 하나이다[7][8][9]. 이러한 연구들을 바탕으로 본 연구에서는 초거대 언어 모델에서 생성된 추론 데이터를 활용하여 감정 분류 작업의 성능을 향상시키는 방법을 제안한다. 이는 감정 분류를 포함한 여러 자연어 처리 작업에 활용될 수 있는 새로운 접근 방식을 제공한다.

## III. The Proposed Scheme

본 연구에서 초거대 언어 모델(ChatGPT)에서 생성된 추론 데이터를 활용하여 감정 분류 성능을 향상시키는 제안 방안은 크게 두 가지 단계로 나뉜다.

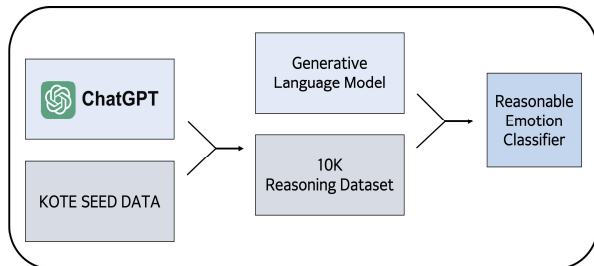


Fig. 1. Proposed Method

실험에서 ChatGPT 모델에 입력되는 KOTE(Korean Online That-gul Emotions) 데이터셋과 해당 데이터셋에 레이블링된 감정 라벨을 사용하였다[10]. KOTE 데이터셋은 다양한 플랫폼에서 수집한 데이터로 구성되어 있으며, 44개의 정서로 레이블링 되어있다. KOTE 데이터셋 50k 중 10k의 데이터를 랜덤 샘플링하여 감정 분류 작업을 위한 Seed Data로 이용하였다.

먼저 첫 번째 단계에서, 초거대 언어 모델인 ChatGPT를 활용하여, 추론 데이터셋(Reasoning Dataset)을 생성한다. KOTE Seed Data로부터 모델에 주어진 입력(Prompt)을 바탕으로 추론하는 과정을 거쳐 사람과 같은 추론 과정을 담은 텍스트 데이터를 얻는 것이다.

## Reasoning Generate Prompting

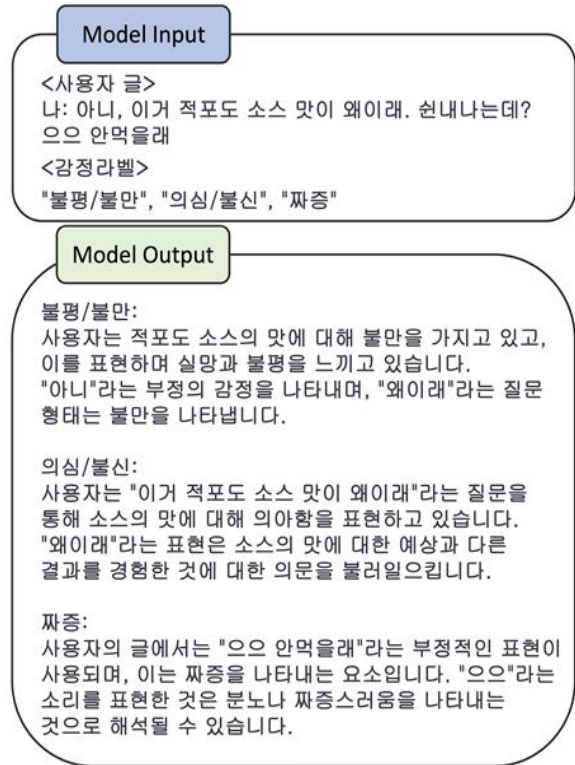


Fig. 2. Reasoning Generate Prompting

Fig. 2는 ChatGPT 모델에 입력되는 KOTE 데이터셋과 해당 텍스트에 대한 감정 라벨이며, ChatGPT 모델을 거쳐 생성된 추론 데이터셋이다.

두 번째 단계에서는 Fig. 2와 같이 생성된 추론 데이터셋을 생성 언어 모델(Generative Language Model; Polyglot-ko-5.8b)에 적용하여 감정 분류 모델 훈련을 진행하였다.

## IV. Experiment

실험에서 ChatGPT 모델을 통해 10k의 데이터셋을 생성하는 비용으로 약 \$45가 소요되었다. 이렇게 ChatGPT 모델을 통해 생성된 추론 데이터셋을 활용하여 감정 분류 모델을 훈련하고, 이를 통해 기존의 데이터셋만을 사용한 모델보다 더 뛰어난 성능을 달성하였다. 감정 분류 작업 실험 결과는 Fig. 3과 같다.

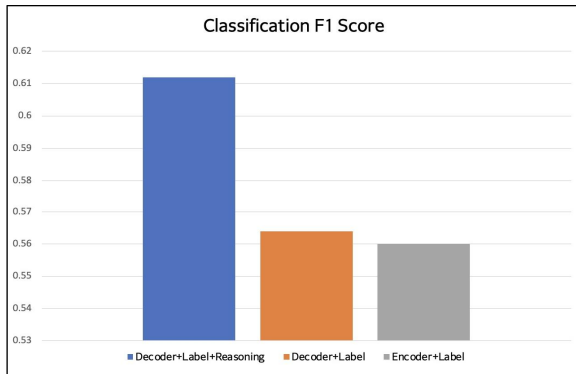


Fig. 3. Classification F1 Score

Fig 3에서 그래프 우측에 위치한 회색 막대는 기존 KOTE 논문에서 사용한 Kc ELECTRA 모델을 미세조정된 F1 Score 값이다[10]. 그래프의 중앙과 좌측의 위치한 막대는 Polyglot-ko-5.8b 모델을 미세조정된 F1 Score 값이다. 그래프 좌측에 위치한 파란색 막대는 ChatGPT로 생성한 추론 데이터셋을 훈련한 모델의 F1 Score 값으로, 막대 그래프를 비교하였을 때 성능이 제일 뛰어난 것을 확인할 수 있다.

이를 통해 추론 데이터셋이 감정 분류 작업에서의 성능을 향상시킨다는 것을 알 수 있으며, 초거대 언어 모델로부터 생성한 추론 데이터셋의 활용 가능성이 높음을 보여준다.

## V. Conclusions

본 연구에서는 초거대 언어 모델로부터 생성된 추론 데이터를 활용하여 감정 분류 성능을 향상시키는 방안을 제시하였다. 고비용의 레이블된 데이터셋을 생성하지 않고 저비용으로 ChatGPT로부터 추론 데이터셋을 생성하여 이를 감정 분류 작업에 적용하였으며, 그 결과 기존의 감정 분류 모델에 비해 향상된 성능을 보였다. 본 논문의 연구 결과는 초거대 언어 모델로부터 저비용으로 생성된 추론 데이터셋이 감정 분류 작업에 있어서 큰 가치를 가질 수 있음을 보여준다.

더 나아가 본 연구는 감정 분류 작업을 넘어서 다른 자연어처리 작업에도 추론 데이터셋이 유용하게 활용될 수 있음을 제안하며, 자연어 이해와 처리를 더욱 정교화하기 위한 새로운 방법론을 시사한다.

## ACKNOWLEDGEMENT

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

## REFERENCES

- [1] ChatGPT (Mar 14 version) [Large language model], <https://chat.openai.com/chat>.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.", NAACL-HLT 4171-4186, 2019.
- [3] Radford, A, Narasimhan, K, Salimans, T, and Sutskever, I, "Improving language understanding by generative pre-training", 2018. <https://lmsys.org/blog/2023-03-30-vicuna/>
- [4] A. Virtanen, J. Kanerva, R. Ilo, J. Luoma, J. Luotolahti, T. Salakoski, F. Ginter, and S. Pyysalo, "Multilingual is not enough: Bert for finnish," arXiv preprint arXiv: 1912.07076, 2019.
- [5] Demszky, D., Movshovitz-Attias, D., Ko, J., Cowen, A., Ne-made, G., and Ravi, S, "GoEmotions: A Dataset of Fine-Grained Emotions", In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 4040-4054, 2020.
- [6] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Ed Chi, Quoc Le, and Denny Zhou, "Chain of thought prompting elicits reasoning in large language models", arXiv preprint arXiv:2201.11903, 2022.
- [7] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, Tatsunori B. Hashimoto. "Alpaca: A Strong, Replicable Instruction-Following Model", 2023.
- [8] Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, Enhong Chen, "A Survey on Multimodal Large Language Models", arXiv preprint arXiv:2306.13549", 2023.
- [9] OpenAI, "GPT-4 Technical Report", arXiv preprint arXiv:2303.08774, 2023.
- [10] Duyoung Jeon, Junho Lee, Cheongtag Kim, "User Guide for KOTE: Korean Online Comments Emotions Dataset", arXiv preprint arXiv:2205.05300Model", 2022.