

# 소셜미디어 텍스트의 슬랭 처리 기반 계층적 감정 분류와 텍스트 감정 맞춤형 이모지 생성 기법

송윤경\*, 임소희\*\*, 김건우\*\*\*

## Hierarchical Emotion Classification for Social Media Texts with Slang Handling and Emotion-Aware Custom Emoji Generation

Yun-Gyeong Song\*, So-Hee Lim\*\*, and Gun-Woo Kim\*\*\*

본 논문은 교육부와 경상남도의 재원으로 지원을 받아 수행된 경상남도 지역혁신중심 대학지원체계(RISE) 연구결과 및 2025년도 산업통상자원부 및 한국산업기술기획평가원(KEIT)의 연구비 지원(RS-2025-02633048)에 의한 연구결과임

### 요약

본 연구는 소셜미디어 텍스트의 정밀한 감정 분석을 반영한 사용자 맞춤형 이모지 생성 방법을 제안한다. GoEmotions를 7개 상위 감정군과 27개 하위 감정군으로 구성된 2단계 계층 구조로 재구성하고, 공유 인코더와 그룹별 헤드를 갖춘 BERT 기반 coarse-to-fine 분류기를 설계하였다. 또한 OED(Oxford English Dictionary)를 기반으로 슬랭 사전을 구축하여 비표준 표현이 포함된 문장의 감정 분류 성능을 향상시켰다. DALL·E 3에 S1(텍스트), S2(텍스트, 감정), S3(텍스트, 감정, 시각 묘사) 프롬프트 전략을 적용하여 텍스트와 감정을 더 충실히 반영하는 최적의 프롬프트를 선정하였다. 실험 결과, 슬랭 정규화를 적용한 계층 분류 모델이 가장 높은 F1-score를 기록했으며, S3 전략이 CLIP-Sentiment, CLIP-Text에서 높은 성능을 기록하였다. 아울러 IS의 저하 없이 이미지 품질과 다양성이 유지되었음을 확인하였다.

### Abstract

This paper proposes a user-tailored emoji generation method grounded in fine-grained emotion analysis of social-media text. We reorganize GoEmotions into a two-level hierarchy (7 superclasses, 27 subclasses) and design a BERT-based coarse-to-fine classifier with a shared encoder and grouped heads. An OED-based slang dictionary with normalization improves robustness to nonstandard expressions. We use the classifier outputs to condition DALL·E 3 via three prompt strategies—S1 (text), S2 (text, emotion), and S3 (text, emotion, Description of emotion). In our experiments, the hierarchical model with slang normalization produced the best F1, and S3 achieved the highest CLIP-Sentiment and CLIP-Text, with no IS degradation, while maintaining image quality and diversity.

### Keywords

hierarchical emotion classification, coarse-to-fine, emoji generation, prompt engineering

\* 경상국립대학교 AI융합공학과  
- ORCID: <https://orcid.org/0009-0008-6692-1925>  
\*\* 경상국립대학교 컴퓨터공학과  
- ORCID: <https://orcid.org/0009-0003-1095-8351>  
\*\*\* 경상국립대학교 컴퓨터공학과 교수(교신저자)  
- ORCID: <https://orcid.org/0000-0001-5643-4797>

· Received: Aug. 29, 2025, Revised: Sep. 22, 2025, Accepted: Sep. 25, 2025  
· Corresponding Author: Gun-Woo Kim  
Dept. of Computer Science and Engineering, College of IT Engineering,  
Gyeongsang National University, Jinju, Korea  
Tel.: +82-55-772-3323, Email: [gunwoo.kim@gnu.ac.kr](mailto:gunwoo.kim@gnu.ac.kr)

## I. 서론

소셜미디어 플랫폼의 확산으로 전 세계적으로 소셜 네트워크 서비스 이용자가 급증하고 있다. Kepios의 글로벌 오버뷰 보고서에 따르면, 2024년 기준 소셜미디어 이용자는 약 50억 4천만 명으로, 전년 대비 2억 6천6백만 명이 증가한 수치이다[1]. 이러한 양적 성장과 더불어, 텍스트 기반의 온라인 소통에서 이모지는 핵심적인 표현 수단으로 자리 잡았다. 사용자들의 적극적인 활용에 힘입어 새로운 이모지가 지속적으로 추가되고 있다. 이모지는 비언어적 소통 도구로서 텍스트에 내포된 감정을 명확히 전달하고 그 강도를 조절하는 역할을 수행한다[2]. 또한, 시각적 정보를 통해 텍스트 기반 소통에서 발생할 수 있는 의미의 모호성을 줄이고, 소통의 간결성을 높인다[3]. 이러한 기능적 장점으로 인해 이모지의 활용도는 꾸준히 증가해왔으며, 2025년 기준 약 3,790종의 이모지가 통용되고 있다.

한편, 이모지 수가 급증하면서 사용자의 텍스트와 감정에 부합하는 최적의 이모지를 신속히 선택하기가 점점 어려워지고 있다. 이에 단순 추천을 넘어 텍스트와 감정을 정밀하게 반영하는 개인화 이모지 생성 기술에 대한 요구가 높아지고 있다.

관련 연구 동향을 보면 감정 분석은 어휘 사전 기반과 전통 기계학습에서 CNN, RNN을 거쳐 최근 Transformer 기반의 미세 감정 분류로 발전해 왔다[4]-[6]. 그러나 소셜미디어 텍스트는 축약어·신조어·오타자 등 언어적 변이가 빈번하고 문장이 짧아, 정제 코퍼스 중심으로 학습된 모델이 실제 환경을 충분히 반영하기 어렵다는 한계가 보고되어 왔다[7]. 또한 세분 감정이 많아질수록 클래스 불균형과 의미적 인접성으로 인한 혼동이 커져, 단일 헤드로 모든 감정을 직접 분류하는 방식은 안정적인 결정 경계 학습에 제약이 있다.

이모지 생성 연구는 템플릿 조합과 아이콘 스타일 전이, GAN 기반 생성 등을 거쳐, 최근에는 텍스트-투-이미지 확산 모델을 활용해 자연어 프롬프트로 이모지를 합성하는 방향으로 빠르게 전개되고 있다[8][9]. 다만 감정 라벨만을 조건으로 한 단순 프롬프트가 여전히 많고, 프롬프트 설계가 텍스트

정합성, 감정 정합성, 품질·다양성에 미치는 영향을 정량적으로 규명한 연구는 드물다.

앞서 언급한 한계를 극복하기 위해, 본 연구는 세 가지 접근법을 제시한다. 첫째, 사용자의 다양한 감정 인식을 위해 공개 데이터셋 GoEmotions를 기반으로 27개 하위 감정군을 7개 상위 감정군으로 계층화하고, 상위에서 하위로 의미 공간을 단계적으로 축소하는 coarse-to-fine 분류 구조를 제안한다. 둘째, 소셜미디어 텍스트의 정확한 의미 해석을 위해 OED(Oxford English Dictionary)기반 슬랭 사전을 구축하고 정규화 절차를 도입한다. 셋째, 텍스트의 미묘한 감정까지 포착하는 프롬프트 엔지니어링을 통해 DALL-E 3[8] 기반 생성 모델로 사용자 맞춤형 이모지 생성 방법을 제시하고, CLIP-Sentiment, CLIP-Text, IS를 통해 의미·감정 정합성과 품질·다양성의 균형을 정량 평가한다. 본 연구는 정적인 추천 방식에서 벗어나, 사용자의 텍스트와 감정에 동적으로 반응하는 개인화된 이모지 생성 방법을 제시했다는 점에서 학문적·실용적 의의를 지닌다.

본 논문의 구성은 다음과 같다. 2장에서는 감정 분석과 이모지 추천 및 생성 관련 연구를 검토하고, 3장에서는 데이터 전처리 및 본 연구의 방법론을 상세히 설명한다. 4장에서는 실험 데이터와 설정, 정량·정성 평가 결과를 제시하며, 마지막으로 5장에서는 결론 및 향후 연구 방향을 제안한다.

## II. 관련 연구

### 2.1 감정 분석 연구

초기 감정 분석 연구는 주로 기쁨, 분노, 슬픔 등 소수의 기본 감정 범주를 대상으로 단일 레이블 분류를 수행하는 데 집중되었다. 그러나 실제 소셜미디어 플랫폼 텍스트는 포괄적인 감정으로 설명하기 어려운 세부적인 감정을 내포하는 경우가 많아, 이 같은 접근법은 복잡한 감정 상태를 포착하는 데 명백한 한계를 보인다. 이러한 한계를 극복하고자, 최근에는 다수의 감정 레이블을 포함하는 대규모 데이터셋이 활발히 구축되고 있다[10][11].

대표적인 예로 구글의 GoEmotions[10]는 27개의

세분화된 감정 범주를 정의하고, 하나의 텍스트에 다수의 감정 레이블을 부여함으로써 복합적인 감정 표현을 모델링하고자 했다. 하지만 감정 범주가 지나치게 세분화될 경우, 클래스 간 의미적 유사성 및 데이터 불균형 문제가 심화되어 모델의 학습을 저해하고 분류 성능의 저하를 야기할 수 있다. 특히, 비정형적인 슬랭 표현과 문맥적 모호성이 짙은 소셜미디어 텍스트의 특성은 이러한 문제점을 더욱 가중시키는 요인으로 작용한다[12].

## 2.2 이모지 추천 및 생성 연구

감정 분석 기술은 사용자의 감정을 시각적으로 표현하는 이모지 추천 시스템에 활발히 응용되고 있다. 해당 기술은 소셜미디어의 사용자 경험 향상부터 감성 챗봇, 정서 피드백 시스템에 이르기까지 폭넓은 잠재적 가치를 지닌다. 그러나 기존 연구들은 대부분 감정 레이블과 이모지를 정적으로 매핑하거나, 사전에 정의된 목록 내에서 선택하는 방식을 채택하고 있다[4][13]. 이러한 접근법은 사용자의 복잡한 감정이나 문맥의 뉘앙스를 반영하는 데 근본적인 한계를 가진다.

이러한 방법은 구현이 간단하고, 기본적인 감정 전달에 효과적이라는 장점이 있으나, 실제 사용자의 미묘한 감정 뉘앙스를 충분히 반영하기에는 한계가 존재한다. 기존 정적 매핑 기반 시스템은 이를 구체적으로 구분하지 못한다.

## 2.3 GPT 및 대규모 언어모델

최근 대규모 언어 모델(LLM, Large Language Model)의 등장은 자연어 처리 패러다임에 큰 변화를 가져왔으며, 특히 자연어 생성의 표현력과 응용 범위를 비약적으로 확장시켰다[14]. LLM은 감정 표현 생성, 프롬프트 최적화, 텍스트-이미지 연동 등 다양한 태스크에서 기존의 규칙 기반 접근법을 효과적으로 대체 및 보완하고 있다. 특히, 별도의 학습 데이터 없이 또는 소수의 예시만으로 새로운 과업을 수행하는 능력은 LLM이 높은 일반화 성능을 보이는 핵심적인 특징이다.

이러한 LLM의 발전은 DALL-E와 같은 텍스트-투

-이미지 생성 모델의 성능 향상으로 이어졌다. 이 모델들은 복잡한 텍스트 프롬프트를 정교한 시각적 결과물로 변환할 수 있으며, 프롬프트의 구조와 표현에 민감하게 반응하여 결과물의 품질이 결정된다. 본 연구는 이러한 LLM의 few-shot 학습 능력과 텍스트-투-이미지 생성 모델의 잠재력을 결합한다. 구체적으로, few-shot 프롬프트 전략을 활용하여 사용자의 감정과 문맥 정보를 정밀하게 반영하는 이모지 생성 프롬프팅 방법을 제안함으로써, 기존 추천 방식의 표현적 한계를 극복하고 개인화된 이모지를 생성하고자 한다.

## III. 사용자 텍스트 기반 이모지 생성 기법 및 구조

본 장에서는 사용자 텍스트에 내재된 감정을 정밀하게 분석하고, 이를 기반으로 문맥에 최적화된 맞춤형 이모지를 생성하는 통합 프레임워크를 제안한다. 제안하는 프레임워크는 아래 네 가지 핵심 방법으로 구성된다. (1) 계층적 감정 데이터셋 구축, (2) 계층적 감정 분석 모델 설계, (3) 소셜미디어 특화 슬랭 사전 구축, (4) 감정-객체 연계 이모지 생성을 위한 프롬프트 네 가지 단계로 구성된다.

### 3.1 데이터셋

본 연구는 구글이 공개한 대규모 감정 데이터셋 GoEmotions[11]를 기반으로 실험 데이터셋을 재구성하였다. 해당 데이터셋은 약 5만 8천 개의 Reddit 댓글과 27개의 세분화된 감정 클래스로 구성되어, 다차원적인 감정 상태를 분석하려는 본 연구의 목표에 부합한다. 그러나 이처럼 세분화된 감정 체계는 클래스 간 데이터 불균형 및 의미적 모호성을 야기하여 모델의 분류 성능을 저하시키는 잠재적 문제를 내포한다. 이러한 문제를 해결하기 위해, 본 연구에서는 27개의 감정 클래스를 의미적 유사도에 따라 7개의 상위 그룹으로 재분류하는 계층적 구조를 설계하였다. 감정 그룹화의 기준은 Plutchik의 감정 이론[15]이 제시하는 심리학적 근거와 원본 데이터셋의 통계적 상관관계 분석[10] 결과를 종합적으로 고려하여 설정함으로써 방법론적 타당성을 확보하였다.

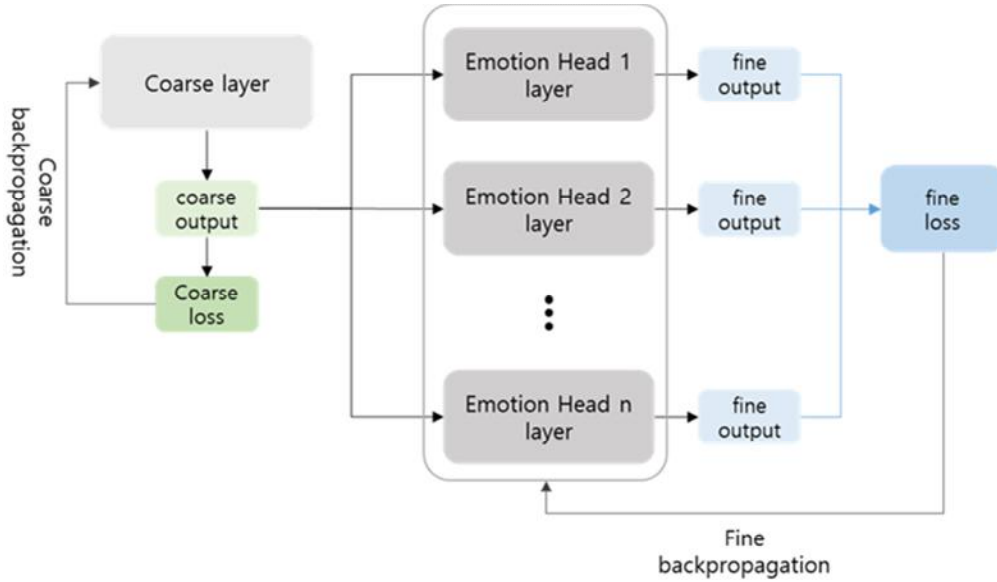


그림 1. 계층적 감정 분류 모델의 coarser-to-fine 학습구조

Fig. 1. Architecture of the hierarchical emotion model with coarse-to-fine training structure

이렇게 재구성된 계층적 레이블 구조는 상위 감정 범주에서 하위의 세부 감정으로 분류를 진행하는 coarse-to-fine 접근법의 적용을 가능하게 한다. 이러한 학습 방식은 세분화된 레이블에서 발생하는 데이터 불균형 문제를 완화하고, 모델 학습의 안정성을 높여 최종적으로 감정 분류의 정확도를 향상시키는 효과를 기대할 수 있다.

### 3.2 슬랭 사전 구축

소셜미디어 플랫폼 기반의 감정 분석에서는 짧은 문장에 등장하는 축약어, 비표준 표현과 같은 인터넷 슬랭 등이 문맥 이해를 어렵게 하여 모델 성능 저하의 원인이 되므로, 의미가 축약된 표현을 정규화하는 과정이 필요하다. 이에 본 연구에서는 소셜미디어 플랫폼 슬랭 사전을 구축하였다. 슬랭 사전 구축에는 OED(Oxford English Dictionary)에서 제공하는 공개 온라인 약어 데이터베이스인 Abbreviations의 Chat 카테고리를 참조하였으며, 약 7,000개 이상의 항목을 수집한 뒤, 숫자 또는 특수 기호가 포함된 항목을 제외하고, 제공되는 사용 빈도 지표가 3 이상인 슬랭만을 선별하여 포함하였다. 사용 빈도가 낮은 항목은 문맥 해석을 왜곡할 위험이 있어 제외하였다.

구축한 사전은 슬랭이 포함된 원문을 의미 보존 상태로 정규화하는 데 활용되며, 이 과정을 통해 문맥 정보가 보완되어 감정 인식 모델의 정확도와 안정성 향상에 기여한다.

### 3.3 계층적 감정 분석 모델 학습 구조 제안

본 연구에서는 재구성된 계층적 데이터셋을 효과적으로 학습하기 위해, 상위 감정과 하위 감정을 순차적으로 고려하는 계층적 분류 모델을 제안한다. 제안하는 모델의 구조는 그림 1과 같다. 먼저 BERT[16] 기반의 공유 인코더가 입력 텍스트의 문맥적 정보를 추출하여 문장 임베딩  $h$ 를 산출한다. 본 연구는 규칙 기반의 추가 조건 벡터는 사용하지 않는다. 생성된 임베딩  $h$ 를 상위 분류기의 입력으로, 상위 레벨 분류기에 전달되어 7개의 주 감정 범주 중 하나로 1차 분류된다. 마지막으로, 해당 분류 결과에 따라 활성화되는 하위 레벨의 감정별 전용 분류기가 27개의 세부 감정 중 최종 감정을 예측하는 coarse-to-fine 구조를 따른다.

$$L_{coarse} = \frac{1}{N} \sum_{i=1}^N CE(\text{softmax}(z_{coarse}^{(i)}), y_{coarse}^{(i)}) \quad (1)$$

식 (1)은 coarse 레이어의 손실함수로, 배치 크기를  $N$ 이라 할 때 상위 7개의 감정군에 대한 손실함수를 의미한다.  $z_{coarse}^{(i)}$ 는 coarse 레이어에서 산출된 로짓 벡터이며  $y_{coarse}^{(i)}$ 는 정답 벡터이다. Coarse 레이어에서 7개 상위 감정군에 대한 단일 라벨 다중분류를 수행한다.

식 (2)는 Fine 레이어의 손실 함수로, 학습 시 상위 감정의 해당 fine 헤드에 대한 손실함수를 의미한다.

$$L_{fine} = \frac{1}{N} \sum_{i=1}^N CE(\text{softmax}(z_{fine}^{(i,g(y_i))}), y_{fine}^{(i,g(y_i))}) \quad (2)$$

$z_{fine}^{(i,g(y_i))}$ 는 coarse 레이어에서 예측된 파인 헤드 레이어에서 산출된 로짓 벡터이며  $y_{fine}^{(i,g(y_i))}$ 는 정답 벡터이다.  $L_{fine}$ 은 해당 감정 파인 헤드로만 학습되며, 코어스 레이어는 학습되지 않는다.

이와 같이 상위 단계에서 의미 공간을 단계적으로 축소하며 분류를 수행함으로써, 학습 안정성과 일반화 성능을 향상시키며, 감정군 간 유사성 관계를 구조적으로 학습할 수 있다.

### 3.4 감정-객체탐지 이모지 생성을 위한 few-Shot 프롬프트

기존 이모지 생성 연구는 주로 감정 라벨 또는 이모지의 외형적 특성만을 기반으로 한 단순 프롬프트를 사용하였으며[17], 이러한 방식은 사용자의 감정이나 텍스트를 충분히 반영하지 못하는 한계가 있다. 이를 해결하기 위해 본 연구는 DALL·E 3를 활용하여 입력 텍스트로부터 감정을 고려하고 이에 적합한 이모지를 생성할 수 있는 few-shot 프롬프트 최적화 전략을 제안한다.

프롬프트는 다음의 설계 원칙을 따른다. 첫째, 불필요한 배경요소나 장식 없이 이모지 스타일을 유지하는 최소한의 표현 사용한다. 둘째, 배경색, 문자, 로고 등 텍스트 삽입 금지와 같은 시각적 제약을 명시한다. 셋째, 소셜미디어 플랫폼에서 통용되는 이모지 형태를 준수한다.

본 연구는 입력 신호의 단계적 강화에 따른 세

가지 변형을 비교한다. 프롬프트 1(S1)은 소셜미디어 텍스트만 제공하여 문장에 내재된 감정과 맥락을 모델이 스스로 추론하도록 유도한다. 프롬프트 2(S2)는 텍스트와 감정 라벨을 함께 제공하여 감정 표현의 명확성과 재현성을 높인다. 프롬프트 3(S3)은 텍스트, 감정라벨, 상위 7개 감정군에 대한 시각 묘사문을 추가하여 감정 표현의 구체성을 강화한다. 텍스트-투-이미지 확산 모델에서 조건을 과도하게 강화하거나, 가이드스 강도를 높이면 정합성은 향상되는 대신 다양성이 감소하는 경향이 보고되어 왔다[18]. 이에 본 연구는 세부 27개 감정 수준의 과도한 시각 묘사 대신 상위 7개 감정군 수준의 시각 묘사로 제한하였다. 이를 기반으로 프롬프트를 구성하였으며 이모지 생성 프롬프트 구조와 감정별 시각 묘사에 대한 정보는 표 1에 제시하였다. 감정별 묘사는 Full Emoji Image Dataset[19]를 참고하여 작성하였으며, 세 프롬프트간 비교를 통해 세부 시각 묘사 제공 수준이 정확성과 감정 표현 반영도에 미치는 영향을 정량적으로 분석한다.

## IV. 실험 및 성능 평가

### 4.1 실험 환경 및 성능 평가

본 연구의 실험은 WSL2 환경의 Ubuntu 18.04 LTS에서 수행되었으며, GPU 가속을 위해 CUDA 12.6과 cuDNN 9.6.0을 설치하였다. 모델 구현에는 Python 3.12.3과 Pytorch 2.18.0을 사용하였으며, 딥러닝 구현을 위해 openai API를 활용하였다. 하드웨어는 Intel Core i513600KF CPU, NVIDIA GeForce RTX4080 GPU, 32GB DDR5 RAM으로 구성되었다.

데이터셋은 GoEmotions 원본을 사용하되 27개 하위 감정을 7개 상위 감정군으로 매핑한 계층 레이블을 부여하였다. 데이터 분할은 K-Fold Cross-Validation으로 수행했으며 K=5로 설정하였다.

모델 및 하이퍼파라미터 설정은 다음과 같다. 공유 인코더로 BERT-base를 사용하고, 상위 7개 감정군 분류를 위한 coarse 헤드와 감정군별 fine 헤드(7개)로 구성된 coarse-to-fine 구조를 적용하였다. 최대 시퀀스 길이는 128, 배치 크기는 32, 옵티마이저는 AdamW, 드롭아웃은 0.1로 설정하였다. 학습률은 인

코더  $2 \times 10^{-5}$ , 헤드  $1 \times 10^{-4}$  로 설정하였으며, 총 30 에폭 학습하고 검증 Macro-F1 기준 early stopping(patience=3)을 사용하였다.

분류 평가는 Precision, Recall, F1-score를 기본 지표로 사용하고, 클래스 불균형을 고려하기 위해 Macro-F1과 Micro-F1을 함께 보고하였다. Precision은 모델이 정답이라고 예측한 사례 중 실제로 정답인 비율을 뜻하며, 거짓 양성의 억제 능력을 정량화한다. Recall은 실제 정답인 사례 중 모델이 놓치지 않고 찾아낸 비율로, 거짓 음성의 억제 및 누락 방지 능력을 나타낸다. F1-score는 Precision과 Recall 사이의 균형을 단일 수치로 요약한 지표다. Macro-F1은 각 클래스의 F1을 동일 가중으로 평균

하여 소수·희소 클래스의 성능까지 균형 있게 반영한 것이며, Micro-F1은 모든 표본을 합산하여 계산함으로써 전체 평균 성능을 나타낸다. 본 연구 맥락에서는 세분 감정 클래스가 많고 분포가 불균형하므로, Macro-F1로 클래스 간 균형 성능을, Micro-F1로 전체 평균 성능을 함께 제시한다. 생성 평가는 분류 결과를 조건으로 생성된 이모지 이미지에 대해 CLIP-Text, CLIP-Sentiment, Inception Score(IS)를 사용하였다. CLIP-Text는 사전학습된 이미지-언어 모델로부터 얻은 이모지 임베딩과 원문 텍스트 임베딩의 유사도를 측정해, 생성 이모지가 입력 문장의 의미를 얼마나 잘 반영하는지를 정량화한다.

표 1. 이모지 생성 프롬프트 구조와 감정별 시각 묘사

Table 1. Emoji generation prompt structure and emotional visual description

Base-prompt	
You are an emoji designer. Your task is to create a single symbolic emoji based on a short text message and its associated emotion. Design only one emoji. The image must include only one flat, round emoji face, centered in the frame. Do not include any other emoji, symbols, logos, or text. Use a plain white background. The emoji must be in the style of modern emoji, flat, round, and minimalistic.	
Description of sentiment	
Joy	A smiling yellow face with gentle eyes and a cheerful smile, radiating happiness and warmth, A beaming face with wide eyes and an energetic smile, showing enthusiasm, A face with heart-shaped eyes or surrounded by hearts, expressing affection, A laughing face with tears of joy and squinting eyes, representing playful humor.
Trust	A calm relaxed face with closed eyes and a soft smile, expressing peace, A face with teary eyes and a warm smile, showing thankfulness, A confident smiling face, symbolizing dignity and self-assurance, A winking or slightly smiling face, showing agreement, A thoughtful face with raised brows, indicating awareness.
Anticipation	A face savoring something with tongue out, representing longing, A bright smiling face with sparkling eyes, showing hope, A warm smiling face with soft eyes, radiating empathy.
Anger	A red face with furrowed brows and clenched teeth, expressing rage, An unamused or expressionless face, showing irritation, A nauseated or vomiting face, representing repulsion, A face with rolling eyes or a frown, showing rejection.
Fear	A wide-eyed face with raised brows and a gasp, showing sudden alarm, A pale face with cold sweat, indicating anxiety, A screaming face with tense features, expressing panic, Eyes darting sideways with clenched teeth, signaling dread.
Sadness	A face with a teardrop, expressing sorrow, A crying face with streaming tears, showing grief, A pensive face with closed eyes and a soft frown, conveying melancholy, A drooping face with a quivering mouth, indicating despair.
Suprise	A face with wide-open eyes, raised brows, and a gaping mouth, expressing astonishment, A thoughtful face with a tilted head, indicating inquisitiveness, A perplexed or dizzy face, expressing uncertainty.

CLIP-Sentiment는 이모지 임베딩과 감정 라벨 임베딩의 유사도를 측정해, 생성 이모지가 라벨 감정을 얼마나 잘 반영하는지를 정량화한다. IS는 분류기 예측 분포의 특성을 이용해 개별 이미지의 확신도와 전체 집합의 다양성을 동시에 반영하는 지표로, 텍스트와 무관하게 이미지 품질과 다양성을 종합적으로 평가한다.

본 연구는 위 지표들을 통해 제안한 분류 파이프라인이 Macro/Micro-F1, Precision, Recall을 향상시키는지, 프롬프트 설계가 의미·감정 정합성 (CLIP-Text/CLIP-Sentiment)을 높이면서도 품질·다양성(IS)을 저해하지 않는지를 체계적으로 검증하였다.

### 4.2 실험 데이터 및 전처리

표 2. 실험 데이터 세트  
Table 2. Experimental data sets

7-class sentiment	27-class sentiment	
Joy	joy	2,607
	admiration	1,520
	excitement	1,900
	love	2,745
	amusement	3081
Trust	relief	452
	gratitude	3,863
	pride	452
	approval	5,928
	realization	2,867
Anticipation	desire	1,248
	optimism	2,887
	caring	1,988
Anger	anger	2,589
	annoyance	4,443
	disgust	1,704
	disapproval	3,774
Fear	fear	1,048
	nervousness	598
Sadness	sadness	2,193
	disappointment	2,771
	grief	227
	remores	849
	embarrassment	817
Suprise	suprise	1,806
	curiosity	3,267
	confusion	2,471

본 연구는 계층적 감정 분석 모델의 성능을 평가하기 위해 구글의 공개 감정 데이터셋 GoEmotions를 사용하였다. Coarse-to-fine 학습을 위해 27개 세부 감정을 7개 상위 감정군으로 매핑하여 계층 라벨을 구성하였다. 표 2는 7개 상위 감정군(Joy, Trust, Anticipation, Anger, Fear, Sadness, Surprise)과 각 감정군에 속하는 27개 세부 감정의 데이터 분포를 나타낸다. Joy 감정군에는 joy, admiration, excitement, love, amusement가 포함되며, Trust 감정군에는 relief, gratitude, pride, approval, realization이 포함된다. 데이터 전처리로는 소문자 변환, URL 멘션, HTML 태그와 같은 SNS 전용 토큰 처리, 슬랭 사전을 이용한 슬랭 변형, 표제어 변형, 공백 처리를 진행하였다.

### 4.3 트랜스포머 모델 기반 감정 계층 학습

본 연구는 제안한 계층형 감정 분류 모델의 성능 검증하기 위해, BERT 기반의 단일 27-class 분류 모델과 제안한 coarse-to-fine 구조의 계층형 모델 비교하였다. 추가로 슬랭 정규화 전처리 유무에 대한 성능 평가 비교도 함께 진행하였다.

단일 27-class 모델은 입력 문장을 BERT 인코더를 통해 임베딩한 후 바로 28개의 감정 클래스를 예측하는 구조이며, 계층형 모델은 먼저 7-class 상위 감정을 분류한 뒤 해당 감정군에 속하는 27-class 세부 감정을 예측하는 구조이다.

표 3은 단일 27-class 모델(flat-27), 계층형 모델(Hire-27), 슬랭 정규화를 적용한 계층형 모델(Hire-27 with slang)의 F1-micro, F1-macro, Precision, Recall 성능을 비교한 결과이다.

표 3. 계층적 감정 분석 모델 성능 비교  
Table 3. Comparison of hierarchical sentiment analysis model performances

Model	F1-micro	F1-macro	Precision	Recall
flat-27	0.36	0.32	0.40	0.32
Hire-27	0.43	0.40	<b>0.45</b>	0.41
Hire-27 with slang	<b>0.44</b>	<b>0.41</b>	0.44	<b>0.42</b>

표 3의 결과에서, 단일 27-class 모델은 f1-micro 0.36, f1-macro 0.32, precision 0.40, recall 0.32로 전반적으로 낮은 성능을 보였다. 이는 감정 클래스 수가 많고 서로 의미적으로 유사한 클래스 산 혼동이 커 단일 헤드가 세부 감정 간 경계를 안정적으로 학습하기 어려웠기 때문으로 해석된다.

반면 계층형 모델은 f1-micro 0.43, f1-macro 0.40, precision 0.45, recall 0.41로 단일 27-class 모델 보다 모든 지표가 개선되었다. 상위 감정군 예측을 통해 탐색 공간을 단계적으로 축소함에 따라 오분류가 줄어든 효과로 볼 수 있다.

또한 계층형 모델에 정규화를 진행한 모델은 f1-micro 0.44, f1-macro 0.41, precision 0.44, recall 0.42로, 슬랭 정규화가 미세한 추가 향상을 보였다.

#### 4.4 GPT few-shot 학습

본 연구는 대규모 언어 모델로 설계한 세 가지 프롬프트 전략(S1-S3)에 대해 DALL·E 3 기반 이모지 생성 성능을 비교하였다. 성능 평가는 CLIP-TEXT, CLIP\_Sentiment[20], IS(Inception Score) [21]를 사용하였다. 각 지표는 생성 이모지의 의미적 일치도, 감정 표현 충실도, 생성된 이모지의 품질과 다양성을 측정한다.

표 4. 이모지 생성 프롬프트 성능 비교  
Table 4. Comparison of emoji generation prompt performances

Prompt	CLIP-text	CLIP-sentiment	IS
S1	0.62	0.65	0.78
S2	0.61	0.71	<b>0.79</b>
S3	<b>0.62</b>	<b>0.73</b>	0.78

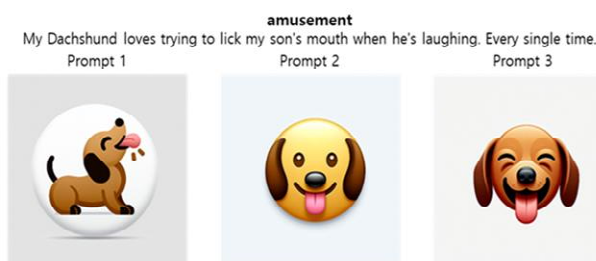


표 4는 세 가지 프롬프트 전략(S1, S2, S3)에 따른 CLIP-Text, CLIP-Sentiment, IS 성능 비교 결과를 나타낸다.

분석 결과 CLIP-Text 점수는 전략 간 유의한 차이가 크지 않았으나 CLIP-Sentiment에서는 S3이 0.73으로 가장 높았다. 이는 감정별 시각 묘사를 추가한 전략이 정서적 일치도를 유의미하게 향상함을 시사한다.

그림 2는 amusement와 curiosity 감정에 대해 각 프롬프트 전략(S1, S2, S3)별로 생성된 이모지 예시를 보여준다. 그림 2에서 확인할 수 있듯이 amusemnet처럼 비교적 명확한 감정의 경우 세 전략 모두에서 감정을 잘 반영한 이모지가 생성되었으나, curiosity처럼 감정이 추상적이거나 해석이 다양한 감정에서는 S1, S2에서 이모지가 감정과 일치하지 않거나 모호한 이모지가 생성되는 경향이 있었다. 반면 S3은 감정 라벨과 함께 구체적인 시각 묘사를 제공함으로써, 이모지의 상징 묘사를 더 명확히 구현하였다.

한편 IS는 전략 간 값이 유사하며, 제안한 프롬프트 강화(S3)가 품질이나 다양성을 저하시키지 않음을 보여준다. 구체적 시각 묘사를 프롬프트에 포함하는 것은 모호한 감정의 해석 정확도를 높이고 감정 전달력을 강화하는데 효과적이다.

#### V. 결론 및 향후 과제

본 연구는 짧은 소셜미디어 텍스트에 대해 coarse-to-fine 계층형 감정 분류기와 슬랭 정규화 사전을 결합하고, 분류 결과를 프롬프트 조건으로 활용하여 DALL·E 3 기반 이모지를 생성하는 방법을 제안하였다.



그림 2. 프롬프트 별 생성 이모지  
Fig. 2. Emojis generated by each prompt

세 가지 구성(단일 27-class 비계층 모델, 계층형 모델, 계층형+슬랭 정규화 모델)을 비교한 결과, 계층형+슬랭 정규화 모델이 F1, Precision, Recall 전반에서 가장 높은 성능을 보였다. 또한 동일한 분류 출력을 조건으로 S1(텍스트), S2(텍스트, 감정), S3(텍스트, 감정, 시각 묘사) 프롬프트를 비교한 결과, S3가 CLIP-Sentiment와 CLIP-Text에서 일관되게 우수했으며 Inception Score(IS)가 유지되어 이미지 품질과 다양성의 저하가 관찰되지 않았다.

본 연구의 의의는 다음과 같다. GoEmotions의 27개 감정을 7개 상위 감정군으로 계층화하고 coarse-to-fine 학습을 적용함으로써, 상위 단계에서 의미 공간을 단계적으로 축소하여 하위 세분 감정을 안정적으로 분류할 수 있었다.

이 과정은 클래스 간 의미적 모호성을 완화하고 학습 안정성·일반화 성능을 높였으며, 감정군 간 유사성 관계를 구조적으로 학습하게 했다. 더불어 슬랭 사전 기반 정규화를 병행하여 비표준·축약 표현으로 인한 언어적 변이를 낮췄다. 이러한 정밀한 계층 분류는 생성 단계에서 프롬프트 조건의 신뢰도를 높여 문장의 핵심 감정을 보다 명확히 전달하고, 추상 감정의 모호성을 줄여 텍스트-감정 정합성을 향상시키면서 이모지의 다양성도 유지하도록 기여하였다.

실제 응용 가능성으로는, 모바일 키보드·메신저에서 실시간 이모지 제안을 통해 추천 선택률을 높이고 선택 소요시간을 단축하여 사용자 편의성을 향상시키는 방안, 소셜 미디어 플랫폼 콘텐츠 작성 보조에서 문장의 톤·추상 감정까지 반영한 권장 이모지 제공으로 사용자 만족도를 높이는 방안이 있다.

향후 연구 방향은 다음과 같다. 첫째, 유사도 기반 가중 손실과 계층 인지 학습을 도입하여 의미적으로 가까운 감정 간 혼동을 줄이고 희소 클래스의 기여도를 높인다. 둘째, 감정-객체 연계 프롬프트를 설계해 생성 단계의 문맥 충실도를 강화하고 감정 뿐만 아니라 관련 객체까지 고려하여 이모지를 생성한다. NER 관계 추출로 얻은 핵심 객체를 상위 7 감정군 시각 묘사 템플릿과 결합하여 조건을 구성하고 이를 토대로 감정과 객체가 반영된 이모지 생성한다.

## References

- [1] We Are Social and Meltwater, "Digital 2024: Global Overview Report", DataReportal report, Jan. 2024. <https://datareportal.com/reports/digital-2024-global-overview-report>.
- [2] M. A. Coyle and C. L. Carmichael, "Perceived Responsiveness in Text Messaging: The Role of Emoji Use", *Computers in Human Behavior*, Vol. 99, pp. 181-189, Oct. 2019. <https://doi.org/10.1016/j.chb.2019.05.023>.
- [3] I. Boutet, M. LeBlanc, J. A. Chamberland, and C. A. Collin, "Emojis Influence Emotional Communication, Social Attributions, and Information Processing", *Computers in Human Behavior*, Vol. 119, No. 106722, Jun. 2021. <https://doi.org/10.1016/j.chb.2021.106722>.
- [4] G. Zhao, Z. Liu, Y. Chao, and X. Qian, "CAPER: Context-Aware Personalized Emoji Recommendation", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 33, No. 9, pp. 3160-3172, Sep. 2021. <https://doi.org/10.1109/TKDE.2020.2966971>.
- [5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need", *arXiv preprint arXiv:1706.03762*, Jun. 2017. <https://doi.org/10.48550/arXiv.1706.03762>.
- [6] T. Mikolov, M. Karafiát, L. Burget, J. Černocký, and S. Khudanpur, "Recurrent Neural Network Based Language Model", *Proc. INTERSPEECH*, Makuhari, Japan, pp. 1045-1048, Sep. 2010. [https://www.isca-archive.org/interspeech\\_2010/mikolov10\\_interspeech.html](https://www.isca-archive.org/interspeech_2010/mikolov10_interspeech.html).
- [7] R. Srivastava and M. P. S. Bhatia, "Challenges with Sentiment Analysis of On-Line Micro-Texts", *International Journal of Intelligent Systems and Applications (IJISA)*, Vol. 9, No. 7, pp. 31-40, Jul. 2017. <https://doi.org/10.5815/ijisa.2017.07.04>.
- [8] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever,

- "Zero-Shot Text-to-Image Generation", Proc. 38th Int. Conf. on Machine Learning (ICML), PMLR, Online, Vol. 139, pp. 8821-8831, Jul. 2021.
- [9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks", arXiv preprint arXiv:1406.2661, Jun. 2014. <https://doi.org/10.48550/arXiv.1406.2661>.
- [10] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A Dataset of Fine-Grained Emotions", arXiv preprint arXiv:2005.00547, May 2020. <https://doi.org/10.48550/arXiv.2005.00547>.
- [11] K. Ding, J. Li, and Y. Zhang, "Hashtags, Emotions, and Comments: A Large-Scale Dataset to Understand Fine-Grained Social Emotions to Online Topics", Proc. 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP), Online, pp. 1376-1382, Nov. 2020. <https://doi.org/10.18653/v1/2020.emnlp-main.106>.
- [12] A. A. Maruf, F. Khanam, M. M. Haque, Z. M. Jiyad, M. F. Mridha, and Z. Aung, "Challenges and Opportunities of Text-Based Emotion Detection: A Survey", IEEE Access, Vol. 12, pp. 18416-18450, Jan. 2024. <https://doi.org/10.1109/ACCESS.2024.3356357>.
- [13] K. R. Narejo, H. Zan, D. Oralbekova, K. P. Dharmani, B. Zhumazhanov, and K. Mukhsina, "Enhancing Emoji-Based Sentiment Classification in Urdu Tweets: Fusion Strategies With Multilingual BERT and Emoji Embeddings", IEEE Access, Vol. 12, pp. 126587-126600, Aug. 2024. <https://doi.org/10.1109/ACCESS.2024.3446897>.
- [14] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding by Generative Pre-Training", OpenAI Technical Report, Jun. 2018. <https://doi.org/10.1177/053901882021004003>.
- [15] R. Plutchik, "A Psychoevolutionary Theory of Emotions", Social Science Information, Vol. 21, No. 4-5, pp. 529-553, Jul. 1982.
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding", Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, pp. 4171-4186, Jun. 2019. <https://doi.org/10.18653/v1/N19-1423>.
- [17] A. Shonenkov, D. Bakshandaeva, D. Dimitrov, and A. Nikolich, "Emojich—Zero-Shot Emoji Generation Using Russian Language: A Technical Report", arXiv preprint arXiv:2112.02448, Dec. 2021. <https://doi.org/10.48550/arXiv.2112.02448>.
- [18] J. Ho and T. Salimans, "Classifier-Free Diffusion Guidance", arXiv preprint arXiv:2207.12598, Jul. 2022. <https://doi.org/10.48550/arXiv.2207.12598>.
- [19] Subinium, "Emoji Image Dataset", Kaggle, 2019. <https://www.kaggle.com/datasets/subinium/emojiiimage-dataset>. [accessed: Aug. 26, 2025]
- [20] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning Transferable Visual Models From Natural Language Supervision", arXiv preprint arXiv:2103.00020, Feb. 2021. <https://doi.org/10.48550/arXiv.2103.00020>.
- [21] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved Techniques for Training GANs", arXiv preprint arXiv:1606.03498, Jun. 2016. <https://doi.org/10.48550/arXiv.1606.03498>.

## 저자소개

송 윤 경 (Yun-Gyeong Song)



2023년 2월 : 경상국립대학교  
컴퓨터공학과(공학사)  
2023년 3월 ~ 현재 :  
경상국립대학교 AI융합공학과  
석사과정  
관심분야 : 인공지능, 데이터 분석,  
생성모델, 멀티모달

임 소 희 (So-Hee Lim)



2020년 3월 ~ 현재 :  
경상국립대학교 컴퓨터공학과  
학사과정  
관심분야 : 인공지능, 데이터 분석,  
자연어 처리

김 건 우 (Gun-Woo Kim)



2006년 12월 : 호주뉴캐슬대학교  
컴퓨터공학과(공학사)  
2007년 9월 : 호주뉴캐슬대학교  
정보공학과(공학석사)  
2017년 8월 : 한양대학교  
컴퓨터공학과(공학박사)  
2021년 9월 ~ 현재 :

경상국립대학교 컴퓨터공학과 부교수  
관심분야 : 인공지능, 시멘틱 헬스케어, 데이터마이닝