

http://dx.doi.org/10.17703/JCCT.2023.9.4.687

JCCT 2023-7-83

## 자연어 처리 딥러닝 모델 감정분석을 통한 감성 콘텐츠 개발 연구

### A Study on the Development of Emotional Content through Natural Language Processing Deep Learning Model Emotion Analysis

이현수\*, 김민하\*\*, 서지원\*\*, 김정이\*\*\*

Hyun-Soo Lee\*, Min-Ha Kim\*\*, Ji-won Seo\*\*, Jung-Yi Kim\*\*\*

**요약** 본 연구는 자연어 처리 딥러닝 모델의 감정분석 정확성을 확인해보고 이를 감성 콘텐츠 개발에 활용하도록 제안한다. GPT-3모델의 개요를 살펴본 후 Aihub에서 제공하는 희곡 대사 데이터 약 6000개를 입력하고 '기쁨', '슬픔', '공포', '분노', '혐오', '놀람', '흥미', '지루함', '통증' 총 9가지 감정 범주로 분류하였다. 이후 자연어 처리 모델 평가 방법인 정확도, 정밀도, 재현율, F1-score 의 평가지표를 활용하여 성능평가를 진행하였다. 감정분석 결과 91% 이상의 정확도를 보였으며 정밀도의 경우 '공포', '통증'이 낮은 수치를 보였다. 재현도의 경우 '슬픔', '분노', '혐오'와 같은 부정적인 감정에서 낮은 수치가 나타났고 특히 '혐오'의 경우 데이터 양의 부족으로 인해 오차가 나타난 것으로 확인된다. 기존 연구의 경우 감정분석을 긍정, 부정, 중립으로 나누는 극성분석에만 주로 사용되어 그 특성상 피드백 단계에서만 사용되는 한계가 있었다. 본 연구는 감정분석을 9가지 범주로 확장하여 기획 단계에서부터 이를 고려한 개발을 통해 게임, 전시, 공연, 관광, 디자인, 에듀테크, 미디어 등에서 감성 콘텐츠 개발에 활용될 수 있음을 제안한다. 후속 연구를 통하여 더욱 다양한 일상 대화들을 추가로 수집하여 감정분석을 진행한다면 더욱 정확한 결과를 얻을 수 있을 것으로 기대된다.

**주요어** : GPT-3, 자연어 처리, 감정분석, 감성콘텐츠

**Abstract** We analyze the accuracy of emotion analysis of natural language processing deep learning model and propose to use it for emotional content development. After looking at the outline of the GPT-3 model, about 6,000 pieces of dialogue data provided by Aihub were input to 9 emotion categories: 'joy', 'sadness', 'fear', 'anger', 'disgust', and 'surprise'. 'interest', 'boredom', and 'pain'. Performance evaluation was conducted using the evaluation indices of accuracy, precision, recall, and F1-score, which are evaluation methods for natural language processing models. As a result of the emotion analysis, the accuracy was over 91%, and in the case of precision, 'fear' and 'pain' showed low values. In the case of reproducibility, a low value was shown in negative emotions, and in the case of 'disgust' in particular, an error appeared due to the lack of data. In the case of previous studies, emotion analysis was mainly used only for polarity analysis divided into positive, negative, and neutral, and there was a limitation in that it was used only in the feedback stage due to its nature. We expand emotion analysis into 9 categories and suggest its use in the development of emotional content considering it from the planning stage. It is expected that more accurate results can be obtained if emotion analysis is performed by additionally collecting more diverse daily conversations through follow-up research.

**Key words** :GPT-3, Natural Language Processing, Emotion Analysis, Emotional Content

\*준회원 성결대학교 미디어소프트웨어학과 학부생 (제1저자) Received: June 25, 2023 / Revised: July 5, 2023  
\*\*준회원 성결대학교 미디어소프트웨어학과 학부생 (참여저자) Accepted: July 10, 2023  
\*\*\*정회원, 성결대학교 미디어소프트웨어학과 조교수 (교신저자) \*\*\*Corresponding Author: ecesss@sungkyul.ac.kr  
접수일: 2023년 6월 25일, 수정완료일: 2023년 7월 5일 Dept. of Media Software, Sungkyul Univ, Korea  
게재확정일: 2023년 7월 10일

## I. 서론

### 1. 연구 배경 및 필요성

최근 자연어 처리 분야에서는 Bert, GPT 등 Transformer 기반의 언어 모델 연구가 활발히 진행되고 있다. 이러한 자연어 처리 모델을 활용하여 시나리오 작성, 리뷰 분석 등 다양한 연구가 이루어지고 있다. 하지만 대부분의 감정분석을 긍정, 부정, 중립만을 활용하는 극성 분석 방법을 사용하고 있다[1,2]. 이를 넘어 더 많은 감정 요소를 사용한다면 활용 범위가 더욱 넓어질 것으로 보인다. 본 연구는 GPT의 대사 감정분석 정확성을 확인해보고 이를 활용하여 감정 콘텐츠 개발에 적용할 수 있음을 제안하고자 한다.

### 2. 연구 목적 및 방법

본 연구에서는 자연어 처리(NLP)에 대해 알아보고 자연어 처리 분야에서 사용이 쉽고 가장 많은 학습 파라미터를 보유한 GPT 모델을 활용하고자 한다. GPT는 대규모 말뭉치를 기반으로 사전 훈련된 모델로 자연어 분석과 생성에 특화되어있다. GPT-3 모델의 개요를 살펴본 후 Aihub에서 제공하는 희곡 대사 데이터를 입력하여 성능평가를 진행하고자 한다.

## II. 이론적 배경 및 선행연구 분석

### 1. 자연어 처리(NLP)

자연어 처리란 인간의 언어 현상을 컴퓨터가 묘사할 수 있도록 연구하는 것을 말한다. 자연어 처리에는 텍스트 데이터를 의미와 맥락에 따라 상관관계를 다차원 벡터로 수치화한 워드 임베딩 기술이 활용된다[3].

자연어 처리에는 RNN 기반 모델, CNN 기반 모델, Transformer 모델이 있다. RNN 기반 모델은 시간의 흐름에 맞춰 변화하는 순서가 있는 데이터를 처리하는데에 적합하다[4]. 이를 변형한 LSTM 모델은 시퀀스 형태의 데이터 처리에 적합하며 다음 나올 단어의 확률을 계산하는데 용이하고 GRU는 LSTM를 구조를 간소화하여 처리속도가 빠르다[5]. CNN 모델은 이미지 처리에서 주로 사용이 되어 합성곱 레이어와 풀링 레이어로 구성된다[4]. Transformer 모델은 2017년 구글에서 발표한 ‘Attention is all you need’에서 처음 등장하였으며, RNN과 달리 양방향 입력 시퀀스와 멀티헤드 어텐션

메커니즘을 활용하여 자연어 처리에 주로 활용되고 연구된다. Transformer 모델은 시간 순서에 따라 입력을 수행했던 RNN과는 달리 입력은 독립적인 값으로 존재하여 위치 인코딩을 포함한다. 셀프 어텐션 구조는 연산 구조상 시작 복잡도가 낮고 병렬화를 통해 적은 컴퓨터 자원으로도 연산이 가능하며 거리가 먼 단어사이에서도 관계성을 찾기 때문에 언어 처리 작업에서 좋은 성능을 보이고 있다[6,7]. Transformer 모델에는 BERT와 GPT등이 있다.

### 2. GPT-3 모델 개요

GPT는 Generative Pre-trained Transformer의 약자로 OpenAI에서 개발한 대규모 언어 모델이다. 3000억 개의 데이터 셋으로 750개의 매개변수를 학습하였고 175억 개의 파라미터로 이루어져 있다[8]. 다음의 그림 1에 나타난 바와 같이 단방향성을 띄며 다음 단어가 무엇인지 맞추는 프리트레인 과정을 거치고 문장 시작부터

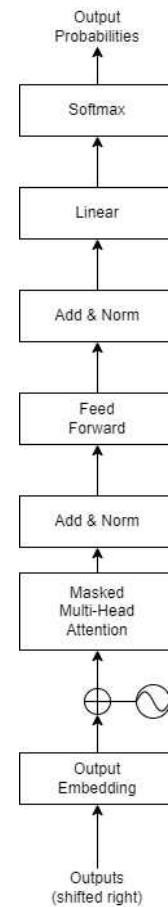


그림 1. GPT 자연어 처리 플로우  
Figure 1. GPT Natural Language Processing Flow

순차적으로 계산한다. 대부분의 Transformer 모델은 양방향성인 반면 GPT의 단방향성은 창작에서 가장 큰 두각을 나타낸다. 본 연구에서는 이러한 특성을 반영하여 대사를 GPT API에 전송하고 대사에 포함된 감정과 행동을 전달받는 감정분석에 사용하고자 한다.

### 3. 감정분석

감성분석이라고도 불리는 감정분석은 텍스트에서 사용자의 감성을 추출하는 것으로 제품이나 장소에 대한 의견의 유사도를 참조할 때 유용하게 사용할 수 있다. 텍스트 속 서비스에 대한 태도나 의견을 분석하는 것이다. 이러한 감정분석을 통해 마케팅 전략, 제품에 대한 피드백 등을 분석할 수 있어서 최근에는 추천 시스템과 감정분석을 결합하여 활용하는 연구가 진행되었다 [9].

하지만 감정분석 연구의 경우 대부분 코퍼스를 활용한 긍정 및 부정에 대한 극성분석만 사용하고 있다 [10, 11]. 극성분석을 수치화하여 통계기법을 활용하여 분석하는 경우가 많은데 그 정확도는 학습 사전이나 모델에 따라 다르게 나타나는데 나상수(2022)의 연구 결과에 따르면 평균 72%의 정확도를 보인다[13]. 극성분석은 만족도를 확인하기 위해 사용되기 때문에 단순히 긍정, 부정, 중립의 값만 요구한다. 본 연구에서는 이러한 극성분석이 갖는 단순성의 한계를 넘어 다양한 감정을 분석하여 더욱 다채로운 감성 콘텐츠로 개발하기 위하여 감성의 범주를 넓혔다. 감정분류는 기업에 관한 뉴스를 수집하여 감정분석을 진행한 김서인(2016)의 연구를 참고하여 Ekman이 정의한 기본 감정 여섯 가지와 활용도가 높은 세 가지 감정 범주를 더하여 ‘기쁨’, ‘슬픔’, ‘공포’, ‘분노’, ‘혐오’, ‘놀람’, ‘흥미’, ‘지루함’, ‘통증’ 총 9 가지 감정 범주[14]를 사용하였다.

## III. 연구방법

### 1. 데이터 수집 및 구축

본 연구에서는 Alhub가 2021년에 구축한 문학작품 낭송, 낭독 음성 데이터 중 회곡 데이터를 수집하여 json 파일을 csv로 변환하여 구성하였다. 불필요한 특수문자와 json 태그를 제거하여 텍스트를 정리하였고 문장 단위로 텍스트를 입력하기 위하여 1000 개의 문장 모음을 5985 개의 대사로 분리하였다. 토큰화 과정에는 GPT API를 사용하여 내부적으로 처리될 것이기 때문에

토큰화 과정은 처리하지 않았다. 그 후 5985 개의 기본적으로 라벨링 된 데이터를 바탕으로 9가지 감정분류로 재분류를 진행하였다. 표 1에서 대사, 실제값, GPT 예측값을 확인할 수 있다.

표 1. 9가지 감정분석 데이터 예시  
 Table 1. 9 Examples of Sentiment Analysis Data

대사	실제값	GPT 예측값
일을! 뭐야? 너 지금 또 징징대려는 거야?	분노	Angry
지옥같이 불안한 나날이었지	공포	Fear
한 달도 채 되기 전에 결혼했다고?	놀람	Surprised

### 2. 모델학습

GPT API를 활용하여 파이썬 코드로 프롬프트를 작성하였고 통신하여 결과값을 도출하였다. 결과값을 정리하면 표 2와 같다.

표 2. GPT 결과값  
 Table 2. GPT Result

감정범주	긍정		부정						
	기쁨	흥미	슬픔	분노	혐오	공포	놀람	통증	지루함
대사 수	1280	1573	829	394	283	429	175	385	629
계	2853		3124						

흥미가 1573개로 가장 많고 기쁨, 슬픔, 지루함, 공포, 통증, 분노, 혐오, 놀람 순으로 감정 범주가 분류되었다.

### 3. 실험 결과 및 분석

다음으로 감정분석의 결과를 비교하기 위해 자연어 처리 모델 평가 방법인 Accuracy, Precision, Recall, F1-score 등의 평가지표를 사용하였다. 이 지표들은 분류분석에서 흔히 사용하는 것들로서 이미 많은 선행연구에서 사용되고 있는 지표들이다.

		실제	
		TRUE	FALSE
판정	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

그림 2. 자연어 처리 모델 평가 분류 매트릭스  
 Figure 2. Natural Language Processing Model Evaluation Classification Metrics

그림 2처럼 실제값을 True 값과 False 값으로 분류하고 판정값을 Positive와 Negative로 나누어 케이스를 분류할 수 있다. True Positive(TP)는 실제 True인 정답을 Positive라고 예측한 정답, False Positive(FP)는 실제 False인 정답을 Positive라고 예측한 오답, False Negative(FN)는 실제 True인 정답을 Negative라고 예측한 오답, True Negative(TN)는 실제 False인 정답을 Negative라고 예측한 정답이다. 다음의 케이스별로 분류한 모델의 성능 평가지표의 수식이 Accuracy, Precision, Recall, F1-score며 수식은 다음과 같다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

정확도는 실제 데이터에서 예측 데이터가 얼마나 같은지를 판단하는 지표로서 가장 직관적인 모델의 성능을 나타낼 수 있는 평가지표이다.

$$Precision = \frac{TP}{TP + FP}$$

정밀도는 판정값이 Positive인 값 중에서 실제값이 True인 것의 비율이다.

$$Recall = \frac{TP}{TP + FN}$$

재현율은 실제값이 True인 것 중에서 판정값이 Positive인 것의 비율이다.

$$F1-score = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}}$$

F1-score은 정밀도와 재현율의 평균이다. 활용도에 따라 정밀도나 재현율의 중요성이 다른데 본 연구의 경우 양성을 음성으로 판단한 경우 사용자나 기획자의 의도대로 수정해도 큰 문제가 발생하지 않지만 음성을 양성으로 잘못 판단한 경우 중요한 정보를 놓칠 수 있기 때문에 재현율보다는 정밀도가 더 중요하다. 이를 바탕으로 각 데이터를 평가한 것을 아래의 표 3에 정리하였다. 소수점 두 자리 이후는 생략하였다.

분석 결과 정확도의 경우 놀람의 경우 0.99로 가장 높고 가장 낮은 통증 값도 0.92로 높은 정확도를 보였다.

하지만 실제 정밀도와 재현도, F1-score의 경우 다르게 나타났다. 정밀도는 공포의 경우 0.65, 통증의 경우 0.47로 낮은 수치를 보였는데 데이터를 살펴보면 나쁜 의도를 성공하였을 때 나타나는 기쁨을 잘 구별하지 못하고 공포로 분류하는 경우와 상대방이 본인을 두려워하는 것을 공포로 분류하는 경우가 있었다. 하지만 대부분의 경우 비슷한 부정적 감정으로 분류하였기 때문에 큰 오차는 없는 것으로 판정된다. 한편 슬픔, 분노, 혐오와 같은 부정적인 감정이 0.57, 0.48, 0.39, 0.3으로 낮은 재현도를 나타냈다. 부정적인 감정의 경우, 특히 혐오는 데이터 양이 적다보니 오차가 나타난 것으로 보인다. 본 연구는 재현도보다는 정밀도가 중요하지만 이를 보완할 수 있는 방법이 필요하다.

표 3. 감성분석 평가지표

Table 3. Sentiment Analysis Evaluation Index

	TP	TN	FP	FN	합계	정확도	정밀도	재현도	F1 score
기쁨	1256	4629	76	24	5985	0.98	0.94	0.98	0.96
슬픔	476	5150	6	353	5985	0.94	0.98	0.57	0.72
분노	193	5584	7	201	5985	0.96	0.96	0.48	0.64
혐오	87	5698	4	196	5985	0.96	0.95	0.30	0.46
공포	419	5340	216	10	5985	0.96	0.65	0.97	0.78
놀람	145	5805	5	30	5985	0.99	0.96	0.82	0.89
흥미	1534	4312	100	39	5985	0.97	0.93	0.97	0.95
통증	372	5193	407	13	5985	0.92	0.47	0.96	0.63
지루함	595	5269	87	34	5985	0.97	0.87	0.94	0.90

#### 4. 감성분석을 통한 감성 콘텐츠 개발 제안

자연어 처리 딥러닝 모델이 분석한 9가지 감성분석은 여러 감성 콘텐츠에서 사용될 수 있다.

콘텐츠 산업에서는 기획 단계에서부터 감성기술을 접목시켜 감성 콘텐츠를 창작할 수 있을 것이다. 게임 분야에서는 NPC의 대사들의 감정을 분석하여 해당 대사의 감정과 일치하는 표정들을 자동으로 처리할 수 있고 채팅 기능에서 채팅의 감정을 분석하여 프로필의 표정을 변화시킬 수 있을 것이다. 전시 분야의 경우 박물관의 다양한 설명이나 역사적 인물들의 명언에서 나타나는 감정들을 분석하여 관련한 이미지들을 활용하는데 도움이 될 수 있다. 공연 분야의 경우 뮤지컬, 연극에서 나타나는 대사의 감정을 분석하여 감정에 맞는 조명들을 자동화할 수 있다. 관광 분야의 경우 관광객들의

리뷰에 나타나는 감정들을 분석하여 활용할 수 있고 디자인에서는 캐릭터의 표정을 자동화 할 수 있을 것이다. 에듀테크에서는 학생들의 글에서 나타나는 다양한 감정들을 분석하여 활용할 수 있고 1인 미디어 분야에서 웹드라마나 버츄얼 아바타에 감정분석을 활용할 수 있을 것으로 보인다. 이러한 예시 외에도 감정분석은 다양한 감성 콘텐츠에서 활용할 수 있을 것으로 기대된다. 하지만 연구 결과 재현도, 정밀도, F1-score가 낮은 부정적 수치의 경우 다소 오차가 있을 수 있으므로 본 연구에서는 개발자와 사용자가 판정된 감정을 변경할 수 있도록 개발할 것을 제안한다.

#### IV. 결 론

본 연구는 GPT의 대사 감정분석 정확성을 확인해보고 이를 웹드라마 영상 제작에 활용할 수 있는 도구를 제안하였다. GPT-3모델의 개요를 살펴본 후 Aihub에서 제공하는 희곡 대사 데이터를 입력하여 성능평가를 진행하였다. 수집된 대사가 희곡 대본이기 때문에 다소 평소에 사용하지 않는 문체들이 포함된 점과 문장의 개수가 일정하지 않은 점, 데이터가 약 6000개로 충분하지 않은 점에서 정밀한 정확도를 나타냈다고 볼 수 없다. 또한 감정분석 특성상 주관적 개입이 있다는 점에서 한계가 있다.

그럼에도 기존의 감정분석을 극성분석 위주로 진행되던 연구를 넘어 9가지 감정으로 분석한 점에서 콘텐츠 개발의 마무리 단계인 피드백 단계에 활용된 극성분석의 한계를 넘어 기획 단계부터 이를 활용할 수 있도록 그 범위를 넓히는 데에 의미가 있다. 또한 각 감정분석의 정확도가 91% 이상인 점에서 상당한 신뢰도를 제공한다. 추후 연구를 통하여 더욱 다양한 일상 대화들을 추가로 수집하여 많은 표본을 나타낸다면 감정분석을 활용하여 감성콘텐츠를 개발하는 데에 도움이 될 것으로 기대한다. 더 나아가 감성콘텐츠를 기획 단계에서부터 적용한 사례들을 통해 보완해나갈 수 있을 것이다.

#### References

- [1] D. H. Jo, I. Jeon and M. K. Moon. "Scenario Generation Assistance System Using GPT-3," *Proceedings of the Korean Society of Computer Information Conference(KSCI)*, Vol.30, No.2, pp. 503-504. July 2022.
- [2] W. S. Sim, Y. H. Moon and H. B. Seo, "Improvement of Online-Shopping Mall Review System using Natural Language Processing," *Electrical Machinery & Energy Conversion Systems Society Annual Spring Conference (KIEE)*, pp.217-218, 2022.
- [3] J. Gaten, K. Sagae and V. Ustun, "Combining distributed vector representations for words," *In Proceedings of the 1st workshop on vector space modeling for natural language processing (NAACL-HLT)*, pp. 95-101, June 2015. DOI: 10.3115/v1/W15-1513
- [4] A. Karpathy, J. Johnson and L. Fei-Fei, "Visualizing and Understanding Recurrent Networks," *arXiv:1506.02078v2*, 2015. DOI: 10.48550/arXiv.1506.02078
- [5] K. H. Cho, B. van Merriënboer and C. Gulcehre, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," *Conference on Empirical Methods in Natural Language Processing(EMNLP)*, pp.1724 - 1734, 2014. DOI: 10.3115/v1/D14-1179
- [6] P. Ekman, "An argument for basic emotions. Cognition and Emotion 6," *Cognition and Emotion*, Vol. 3, No. 4, pp. 169-200, Oct 1992. DOI: /10.1080/02699939208411068
- [7] A. Vaswani, N. Shazeer and N. Parmar, "Attention is all you need," *Advances in Neural Information Processing Systems 30*, pp. 6000-6010, Dec 2017.
- [8] Z. Hwang, "Understanding and Trends of a Transformer Deep Learning Model in the Liberal Arts Education," *The Journal of Liberal Arts (HILAS)*, Vol.9, No.2, pp. 145-156, 2022. DOI: 10.55450/hilas.2022.9.2.145
- [9] T. Brown, "Language models are few-shot learners," *Advances in Neural Information Processing Systems 33(NeurIPS)*, pp. 1877-1901, 2020. DOI: 10.48550/arXiv.2005.14165
- [10] H. G. Jeon, K. C. Jung and K. A. Kwon, "Machine Learning Language Model Implementation Using Literary Texts," *The Journal of the Convergence on Culture Technology(JCCT)*, Vol.7 No.2, pp. 427-436, 2020. DOI : 10.17703/JCCT.2020.6.2.449
- [11] Jae-Kyung Sung, Yung-Bok Kim and Yong-Guk Kim, "Deep learning-based Multilingual Sentimental Analysis using English Review Data," *The Journal of the Institute of Internet (JIIB)*, Broadcasting and Communication, Vol.

- 19, No.3, pp.9-15, 2019. DOI : 10.7236/JIIBC.2019.19.3.9
- [12]Hyorim Shin, Junho Choi, “Analysis of User Reviews for Webtoon Applications Using Text Mining,” *The Journal of the Convergence on Culture Technology(JCCT)*, Vol.8 No.4, pp. 457-468, 2022. DOI : 10.17703/JCCT.2022.8.4.457
- [13]S. S. Na, J. Y. Kang and S. J. Rlee, “Implementation of Age Prediction Model for Elementary and Secondary Learners’ Writings Using Transformer Language Model,” *Journal of CheongRam Korean Language Education*, No.90, pp. 51-96, 2022. DOI : 10.26589/jockle..90.202211.51
- [14]S. I. Kim, D. S. Kim and J. W. Kim, “Public Sentiment Analysis of Korean Top-10 Companies : Big Data Approach Using Multi-categorical Sentiment Lexicon,” *Journal of Intelligence and Information System(JIIS)*, Vol.22, No.3, pp.45-69, 2016. DOI : 10.13088/jiis.2016.22.3.045