

멀티모달 감정 인식 AI 기술을 이용한 우울증 예방 플랫폼 구축

장현빈¹, 조의현², 권수연³, 임신민⁴, 조세린⁵, 나정은⁶

¹연세대학교 생명시스템대학 생화학과 학부생

²연세대학교 공과대학 산업공학과 학부생

³연세대학교 문과대학 노어노문학과 학부생

⁴연세대학교 언더우드국제대학 생명과학공학과 학부생

⁵연세대학교 생명시스템대학 생명공학과 학부생

⁶연세대학교 학부대학 교수

kori834@yonsei.ac.kr, cuihyun@yonsei.ac.kr, suyeon08@yonsei.ac.kr,

smlim02@yonsei.ac.kr, lynn0804@yonsei.ac.kr, jenah@yonsei.ac.kr

Development of a Depression Prevention Platform using Multi-modal Emotion Recognition AI Technology

HyunBeen Jang¹, UiHyun Cho², SuYeon Kwon³,

Sun Min Lim⁴, Selin Cho⁵, JeongEun Nah⁶

¹Dept. of Biochemistry, Yonsei University

²Dept. of Industrial Engineering, Yonsei University

³Dept. of Russian Language and Literature, Yonsei University

⁴Dept. of Life Science and Biotechnology, UIC, Yonsei University

⁵Dept. of Biotechnology, Yonsei University

⁶University College, Yonsei University

요 약

본 연구는 사용자의 음성 패턴 분석과 텍스트 분류를 중심으로 이루어지는 한국어 감정 인식 작업을 개선하기 위해 Macaron Net 텍스트 모델의 결과와 MFCC 음성 모델의 결과 가중치 합을 분류하여 최종 감정을 판단하는 기존 82.9%였던 정확도를 텍스트 모델 기준 87.0%, Multi-Modal 모델 기준 88.0%로 개선한 모델을 제안한다. 해당 모델을 우울증 예방 플랫폼의 핵심 모델에 탑재하여 covid-19 팬데믹 이후 사회의 문제점으로 부상한 우울증 문제 해소에 기여 하고자 한다.

1. 서론

우울증은 초기 진단과 대응이 중요함에도, 사회적 편견과 경제적 어려움으로 인해 많은 사람들이 적절한 치료 시기를 놓치고 있다. 특히 우울증 유병률이 covid-19 팬데믹을 기점으로 25% 증가했다는 세계 보건 기구(WHO)의 발표[1]는, 우울증 예방 및 조기 진단을 위하여 개인이 감정을 인지하고 관리할 수 있는 효과적인 수단의 필요성을 촉구한다.

기존 감정 분석에서는 CNN, RNN, LSTM 등의 자연어 처리 알고리즘을 사용하며, 한국어 감정 분석을 위해서는 KoBERT 모델[2]을 주로 활용한다. 본 연구에서는 한국어 음성 데이터를 바탕으로, 딥러닝 멀티모달 모델을 구현하여 '감정 예측 및 우울증 예방 플랫폼'을 구축한다. 기존 모델에 비해 한국어 음성에 대한 감정 분석 성능을 개선하고, 플랫폼을 앱 형태로 구현하여 그 접근성을 높이고자 한다.

2. 감정 예측 및 우울증 예방 플랫폼 구축

본 연구에서는 한국어 음성을 text로 변환한 후 text 감정 분석을 진행하는 모델과 음성 자체의 감정 분석을 진행하는 모델이 결합된 형태의 Multi-Modal 모델[3]을 구현한다. 특히, text 감정 분석 모델은 Multi-particle Dynamic System과 Macaron Net 구조[4]를 차용하여 그 성능을 높인다.

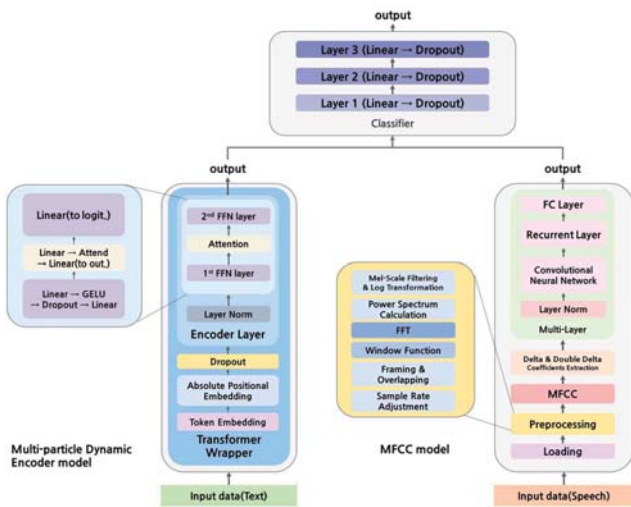
2.1 Text 감정 분석 모델 성능의 개선

한국어 text 감정 분석 모델의 성능 향상을 위해 Multi-particle Dynamic System을 도입하였다. Transformer의 각 layer에서 Euler's method를 통해 수치적으로 근사한 벡터를 처리하여 단어가 문맥에 따라 추상화되는 것을 학습한다. <표 1>은 기존 Bert-based 모델의 성능에 비하여, Multi-particle Dynamic System과 Macaron Net 구조를 차용한 모델의 성능이 5% 이상 향상됨을 보여준다.

<표 1> Multi-particle Dynamic 모델과 기존 Bert-based 모델들의 한국어 Text 감정 분석에 대한 성능 비교

Encoder model	Accuracy	score		model		loss
		F1	Precision Recall	num of Parameter (백만)	size (MB)	
Multi-particle Dynamic & Macaron Net	0.870	0.870	0.885 0.870	46.0	184.04	0.415
		0.825				
		0.843				
KoBERT	0.829	0.829	0.829	92.2	368.77	0.516
		0.769				
		0.787				
DistillBERT	0.772	0.772	0.772	135	541.32	0.632
		0.794				
		0.813				
multilingual BERT	0.798	0.798	0.798	177	711.44	0.586
		0.798				

2.2 한국어 음성 감정 분석 Multi-Modal 모델 (Macaron Net Encoder & MFCC) 구현



(그림 1) 한국어 음성 감정 분석 Multi-Modal 모델 구조도

Multi-modal 모델은 (그림 1)과 같은 구조로 이루어진다. 해당 모델은 개선된 Encoder 모델의 output과 MFCC[5] 기반 모델의 output을 결합한 결과를 fc layer에 통과하는 방식으로 감정 분류를 수행한다. 감정 class에 대한 분류 성능은 <표 2>와 같다.

<표 2> Multi-Modal 모델의 감정 class 별 분류 성능

	score			support
	F1	Precision	Recall	
기쁨	0.92	0.88	0.95	435
슬픔	0.91	0.96	0.87	177
분노	0.82	0.87	0.77	311
상처	0.94	0.93	0.94	169
불안	0.85	0.86	0.83	238
당황	0.88	0.88	0.88	292
중립	0.87	0.84	0.90	316
Total	Accuracy	0.88	Loss	0.401

2.3 Multi-Modal 모델(Macaron Net Encoder & MFCC)을 적용한 감정 예측 앱 구현

(그림 2)는 개발된 애플리케이션 화면으로, 음성 녹음 화면, 미션 화면, 7일 이상 데이터가 누적된 후 사용할 수 있는 감정 분석 화면으로 구성되어 있다. 사용자에게 본인의 상태를 파악하여 휴식을 취하도록 유도하거나, 더 심화 된 단계로 우울감이 진행되는 것을 예방할 수 있도록 설계하였다.



(그림 2) Multi-Modal 모델을 적용한 앱 구동 형태

3. 결론

본 연구에서는 Transformer Encoder 구조에 Macaron Net 구조 및 Multi-particle Dynamic System을 적용하여 기존 KoBERT 및 BERT 기반 한국어 감정 분석 모델보다 5% 이상의 분석 결과를 높였으며, 모델의 용량과 parameter를 줄이는 방식으로 성능을 개선하였다. 또한, 이를 MFCC 모델과 결합하여 성능을 88.0%로 상향 개선한 한국어 감정 분석 Multi-Modal 모델을 구현하였다.

본 논문에서 제안한 모델을 사용자가 접근하기 쉬운 앱 형태로 제작하여, 일상 속에서 사용자들이 자신의 감정을 인지하고 감정을 건강한 방향으로 해소할 수 있는 플랫폼을 구축하였다.

[본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.]

참고문헌

[1] Kupcova, Ida, et al. "Effects of the COVID-19 Pandemic on Mental Health, Anxiety, and Depression." BMC Psychology, U.S. National Library of Medicine, no. 108, 2023.

[2] Young-Jun Lee, and Ho-Jin Choi. "Joint Learning-based KoBERT for Emotion Recognition in Korean." 한국정보과학회 학술발표논문집 (2020): 568-570.

[3] Hazman, Muzhaffar, et al. "Unimodal Intermediate Training for Multimodal Meme Sentiment Classification." ArXiv.org, 1 Aug. 2023, arxiv.org/abs/2308.00528.

[4] Lu, Yiping, et al. "Understanding and Improving Transformer from a Multi-Particle Dynamic System Point of View." ArXiv.org, 6 June 2019, arxiv.org/abs/1906.02762.

[5] Shukla, Rachit. "Keywords Extraction and Sentiment Analysis Using Automatic Speech Recognition." ArXiv.org, 7 Apr. 2020, arxiv.org/abs/2004.04099.