

나이·성별 예측과 멀티모달 감정 인식을 활용한 개인 맞춤형 음악 추천 시스템 개발

Development of a Personalized Music Recommendation System Utilizing Age, Gender Prediction, and Multimodal Emotion Recognition

연구자 : 이석현 (동양미래대학교 자동화공학과), 김해현 (동양미래대학교 자동화공학과)

지도교수 : 황일규 (동양미래대학교 자동화공학과 교수)

요 약

이 프로젝트는 나이와 성별 예측, 멀티모달 감정 인식을 기반으로 개인 맞춤형 음악을 추천하는 시스템을 개발하는 것을 목표로 하였다. 컴퓨터 비전 기술을 활용해 사용자의 얼굴 데이터를 분석하여 나이와 성별을 예측하였으며 텍스트 자연어 처리 및 음성 처리를 결합한 멀티모달 접근 방식을 통해 사용자의 감정을 정밀하게 분석하였다. 이를 바탕으로 YouTube API와 MIDI 알고리즘을 활용하여 사용자의 감정과 인구통계적 특성에 적합한 음악을 추천하거나 직접 작곡하는 기능을 구현하였다.

프로젝트는 고립된 생활이나 일상적 스트레스 속에서 감정적 공감과 정서적 만족을 제공할 수 있는 혁신적인 개인화 솔루션을 제안한다. 결과적으로 이 시스템은 사용자의 감정을 반영한 음악 경험을 통해 삶의 질을 향상시키는데 기여할 것으로 기대된다.

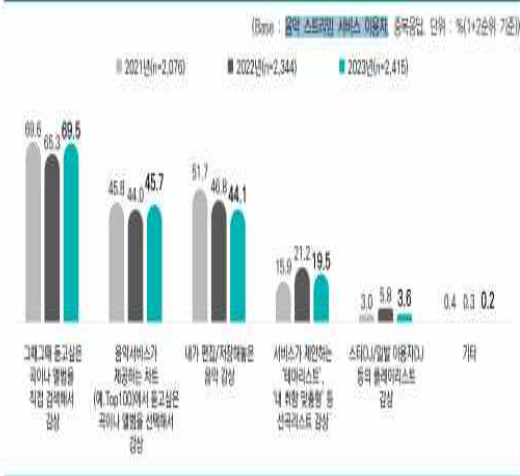
▶ Keyword : 멀티모달 감정 인식 (Multimodal Emotion Recognition), 개인 맞춤형 음악 추천 (Personalized Music Recommendation), 나이와 성별 예측 (Age and Gender Prediction), 컴퓨터 비전 (Computer Vision), 자연어 처리 (Natural Language Processing, NLP), 음성 감정 인식 (Speech Emotion Recognition, SER), MIDI 작곡 (MIDI Composition), 딥러닝 (Deep Learning), 전이 학습 (Transfer Learning), 감정 기반 음악 추천 (Emotion-based Music Recommendation), 인간-컴퓨터 상호작용 (Human-Computer Interaction, HCI)

I. 서론

1-1. 연구목적

현대 사회에서 많은 사람들이 감정과 상황에 맞는 음악을 찾는 데 어려움을 겪고 있습니다. 다양한 음악 플랫폼이 존재하지만 대부분의 추천 시스템은 사용자의 감정 상태나 성별, 나이대 와 같은 개인적 요소를 충분히 고려하지 못하고 있습니다. 이로 인해 음악 추천 결과가 때로는 사용자의 기대에 부합하지 않는 경우가 빈번합니다.

그림 2-2-2 음악 스트리밍 서비스 이용 방법



<Fig. 1> 음악 스트리밍 서비스 이용 방법

한국콘텐츠진흥원이 발표한 '2023 음악 이용 (음악 이용자 실태조사, 2023)자 실태조사'(음악 이용자 실태조사, 2023)에 따르면 음악 스트리밍 서비스 이용자 중 상당수가 자신이 원하는 곡이나 앨범을 직접 검색하여 감상(69.5%)하는 방식을 선호하는 반면 서비스가 제안하는 '테마리스트'나 '내 취향 맞춤형' 선곡리스트를 이용하는 비율은 상대적으로 낮게 나타났습니다(19.5%). 이러한 데이터는 현재의 추천 시스템이 사용자들의 개별적 특성 특히 성별과 나이대, 감정 상

태를 충분히 반영하지 못하고 있음을 시사하며 이는 만족도의 저하로 이어질 수 있음을 보여줍니다.

표 2-8-3 응답자 특성별 인공지능 작곡 음악 감상 의향

(Base : 전체 단위 : %)

구분	사례 수	의향 없음		의향 있음		의향 없음		의향 있음	
		전혀	별로	보통	의향 있음	다소	매우	의향 있음	의향 있음
전체	(3,500)	24.7	7.3	17.4	42.2	33.1	29.7	3.4	
성별	남성 (1,781)	23.5	6.7	16.8	43.1	33.4	29.5	3.9	
	여성 (1,719)	25.9	7.9	18.0	41.2	32.9	30.0	2.9	
연령별	10대 (415)	25.5	7.7	17.8	41.4	33.0	29.2	3.9	
	20대 (591)	28.8	9.6	19.1	37.4	33.8	30.6	3.2	
	30대 (681)	25.8	8.4	17.5	48.2	26.0	23.2	2.8	
	40대 (606)	20.9	6.2	14.8	43.3	35.8	31.1	4.7	
	50대 (677)	23.2	6.1	17.1	42.4	34.4	31.3	3.1	
	60대 (438)	24.7	5.7	18.9	38.1	37.2	34.7	2.5	

<Fig. 2> 인공지능 작곡 음악 감상 의향

한편 같은 조사(음악 이용자 실태조사, 2023)에서 인공지능 작곡 음악 감상에 대한 이용자 의향을 분석한 결과 응답자의 66.7%가 인공지능이 작곡한 음악 감상에 긍정적 의향을 보였으며 특히 50대와 60대 응답자들 사이에서 그 관심이 높게 나타났습니다(각각 67.1%, 72.1%). 이러한 데이터는 인공지능 기술이 음악 추천 시스템에 통합될 경우 사용자 만족도를 크게 향상시킬 수 있는 잠재력이 있음을 시사합니다.

본 연구는 이러한 배경을 바탕으로 나이와 성별 예측, 감정 분석, 그리고 인공지능 작곡 기술을 결합하여 사용자 맞춤형 음악 추천 시스템을 개발하는 데 중점을 두고 있습니다. 이를 통해 사용자들이 개인의 감정과 상황에 부합하는 음악을 더 쉽게 찾을 수 있도록 지원하며, 궁극적으로 음악 소비 경험을 혁신적으로 개선하고자 합니다.

1-2. 연구의 범위와 방법

1. 연구의 범위

본 연구는 음악 추천 시스템의 개인화를 목표로 성별과 나잇대, 감정 상태를 분석하여 사용자의 현재 상황에 맞는 음악을 추천하는 시스템을 설계하는 데 초점을 맞추고 있다. 이를 위해 다음 범위를 설정하였다.

㉠ 컴퓨터 비전 기반 사용자 분석 : 사용자의 얼굴 데이터를 기반으로 성별과 나잇대를 검출하여 음악 선호도와 상태에 대한 개인화된 추천의 기초 자료로 활용한다.

㉡ 멀티모달 감정 분석 : 텍스트(자연어 처리)와 음성 데이터를 통해 감정 상태를 세분화(예: 행복, 슬픔, 화남 등)하여 사용자 정서에 적합한 음악을 추천한다.

㉢ 인공지능 작곡 기술 통합 : AI 작곡 기술을 활용하여 추천된 음악 외에도 사용자의 감정에 맞는 새로운 음악을 생성하고 감상할 수 있는 기능을 제공한다.

㉣ YouTube 음악 추천 시스템 : YouTube API를 활용하여 사용자의 감정과 나이에 따라 적합한 음악을 실시간으로 검색 및 재생하는 기능을 구현한다. 감정 분석 결과를 바탕으로 사용자가 선호할 만한 노래를 추천하며, 재생 후 사용자의 피드백을 통해 추천 알고리즘을 지속적으로 학습시켜 정확도를 높인다. 이는 기존의 음악 스트리밍 서비스가 가진 한계를 보완하고, 개인의 정서와 상황에 맞춘 맞춤형 추천 시스템으로 발전하는 것을 목표로 한다.

㉤ 사용자 만족도 향상 : 한국콘텐츠진흥원의 '2023 음악 이용자 실태조사' 데이터에서 파악된 추천 시스템의 한계점을 보완하고 추천 음악의 감정 적합성을 높이는 것을 목표로 한다.

2. 연구의 방법

본 연구의 설계 및 실행 방법은 다음과 같다.

㉠ 데이터 수집 및 전처리

컴퓨터 비전: OpenCV 및 ResNet 모델을 사용하여 사용자 얼굴 데이터를 분석하고, 성별과 나잇대를 예측한다.

자연어 처리: KCElectra 모델을 활용하여 텍스트 데이터를 기반으로 감정 라벨을 분류한다.

음성 분석: Wav2Vec2 모델을 사용해 음향적 특징(음의 강도, 주파수)을 추출하고 감정을 분석한다.

㉡ 멀티모달 감정 분석

텍스트, 음성 데이터를 통합한 멀티모달 모델을 설계하고 가중치 기법을 사용하여 2 가지 입력 데이터를 결합하고 최종 감정 상태를 분류한다.

㉢ 추천 시스템 설계

멀티모달 감정 분석 결과를 기반으로 사용자의 감정과 성별·나잇대에 적합한 음악을 추천

YouTube API와 AI 작곡 모델을 결합하여 사용자 경험을 개인화 하고 피드백 시스템을 도입해 사용자가 추천 음악에 만족하는지 여부를 기록하고, 이를 통해 감정별 가중치 조정 및 추천 알고리즘을 개선한다.

㉣ 시스템 평가

연구에서 개발한 시스템은 사용자 데이터를 활용하여 성능을 평가하며 만족도 조사와 추천 음악의 적합성 분석을 통해 개선점을 도출

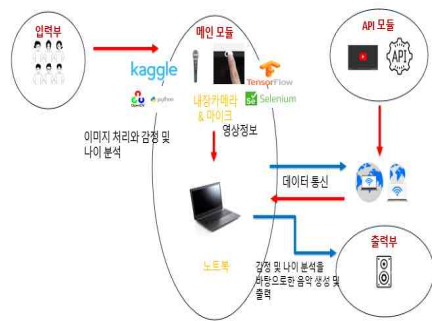
II. 시스템 설계

2-1. 설계 개요

본 연구는 컴퓨터 비전, 자연어 처리, 음성 처리 기술을 융합하여 사용자의 나이와 성별을 예

측하고 텍스트와 음성을 기반으로 사용자의 감정을 분석한 뒤 이를 활용해 개인화된 음악 추천 시스템을 설계하는 데 중점을 두었다. 시스템은 입력부, 메인 모듈, API 모듈, 출력부로 구성되며 각 모듈은 데이터 처리 및 통합의 역할을 담당한다.

아래 그림은 본 연구에서 설계한 시스템의 전체 구조를 나타낸다.



〈Fig.3〉 시스템 구조도

〈Fig.3〉을 보았을 때 시스템은 크게 입력부, 메인 모듈, API 모듈, 출력부로 나뉘는 것을 알 수 있다.

㉠ 입력부 : 사용자의 얼굴 이미지를 통해 나이와 성별을 분석하며 마이크를 통해 음성 데이터를 수집한다. 이를 위해 OpenCV, Python 라이브러리를 활용하여 이미지 및 음성 데이터를 처리한다.

㉡ 메인 모듈 : 입력부에서 수집된 데이터를 Kaggle, TensorFlow, Selenium 등의 기술을 통해 처리한다.

내장 카메라 및 마이크를 사용하여 실시간으로 데이터를 수집하고 분석하며 분석 결과를 노트북으로 전송한다.

㉢ API 모듈 : 분석된 감정 및 나이 데이터를 활용하여 YouTube API와 기타 외부 API를 통해 사용자 맞춤형 음악을 추천한다.

㉤ 출력부 : 분석 결과를 바탕으로 생성된 음악 및 추천 정보를 사용자에게 제공하며 스피커나 디스플레이를 통해 출력한다.

본 시스템은 텍스트, 음성, 이미지 데이터를 융합하여 사용자의 감정을 종합적으로 분석하고 이를 기반으로 개인화된 음악을 추천하는 멀티모달 접근 방식을 채택하였다. 입력 데이터는 실시간으로 처리되며 분석 결과는 즉각적으로 사용자에게 제공될 수 있도록 설계되었다. 이를 위해 OpenCV, TensorFlow, Selenium, Python과 같은 다양한 기술을 활용하여 이미지 및 음성 데이터를 효과적으로 처리하였으며 외부 API를 연동하여 사용자 경험을 극대화하였다. 이러한 설계는 데이터의 통합 분석과 실시간 처리의 장점을 결합하여 높은 정확성과 효율성을 보장한다.

2-2. 주요 기술

본 연구에서는 컴퓨터 비전 기술을 활용하여 사용자의 나이와 성별을 예측하고 자연어 처리와 음성 처리 기술을 결합하여 텍스트와 음성 데이터를 기반으로 사용자의 감정을 분석하는 시스템을 설계하였다.

㉠ 나이 학습

나이 학습은 UTKFace 데이터셋을 기반으로 진행되었으며 나이를 범주화하여 학습 모델의 효율성을 향상시켰다. 예를 들어 0-9세, 10-19세, 20-39세, 40-59세와 같은 연령대로 나뉘어 학습되었다.

VGG16 모델은 깊고 단순한 구조로 인해 이미지 분류 작업에 적합하며 학습 결과 약 85% 이

상의 정확도를 기록하였다.

㉞ 성별 학습

성별 학습에서는 UTKFace 데이터셋과 VGG16 모델을 사용하여 남성과 여성으로 분류하였다.

학습 과정에서 Binary CrossEntropy Loss를 사용하였으며 약 90% 이상의 높은 분류 정확도를 달성하였다. 또한, 과적합 문제를 방지하기 위해 Dropout 레이어를 추가하였으며 조도와 환경 변화에 강건한 성능을 유지하도록 데이터 전처리를 강화하였다.

㉟ 텍스트 자연어 처리

텍스트 데이터를 기반으로 사용자의 감정을 분석하기 위해 KcELECTRA 모델을 활용하였다. 주요 학습 데이터셋으로는 AIHub 감정 데이터셋을 사용하였으며 각 데이터에는 감정 레이블(행복, 슬픔, 화남, 중립, 놀람, 두려움, 혐오)이 포함되어 있다.

전이 학습(Transfer Learning) 기법이 적용되었으며 KcELECTRA는 사전 학습된 한국어 언어 모델로 감정 분석 작업에 특화되도록 Fine-tuning되었다. 이러한 전이 학습 방식은 대규모 일반 텍스트 데이터로 학습된 사전 학습 모델을 감정 분석과 같은 특정 작업에 맞게 미세 조정하여 적은 양의 데이터로도 높은 성능을 달성할 수 있도록 한다.

㊱ 음성 처리

음성 데이터를 활용하여 사용자의 감정을 분석하기 위해 Wav2Vec2 모델을 사용하였다.

주요 데이터셋으로는 AIHub 대화 음성 데이터셋을 사용하였으며 음성 신호에서 MFCC와 같은 음향 특징을 추출하여 모델 입력 데이터로 사용하였다.

Wav2Vec2 모델은 24개의 Transformer 레이

어로 구성되어 있으며 음성 데이터를 고차원 특징으로 변환하여 7가지 감정 상태를 분류하였다.

㊲ 멀티모달 감정 분석

본 연구에서는 텍스트 자연어 처리와 음성 처리 결과를 결합하여 사용자의 감정을 보다 정밀하게 분석하는 멀티모달 감정 분석 모델을 설계하였다. 멀티모달 접근법은 단일 입력 데이터 방식의 한계를 보완하며, 서로 다른 데이터의 상호 보완적 특성을 활용함으로써 감정 분석의 정밀도와 신뢰성을 향상시키는 데 중점을 두었다.

텍스트와 음성 데이터의 결합에는 가중치 방식(Weighted Fusion)을 적용하였으며 텍스트와 음성의 가중치를 8:2로 설정하였다. 이러한 설정은 텍스트 모델(KcELECTRA)이 데이터 안정성과 높은 성능을 보여 감정 분석의 주요 역할을 담당하도록 설계한 데 기인한다. 반면 음성 모델(Wav2Vec2)은 텍스트 모델의 분석 결과를 보완하는 역할로 사용되었다. 이를 통해 텍스트 기반 감정 분석의 강점을 유지하면서도 음성 데이터를 추가적으로 반영하여 감정 추론의 정밀도를 보완하였다.

결합 방식은 다음과 같다. 텍스트와 음성 모델의 출력 결과를 Weighted Fusion 방식으로 통합하여 최종 감정을 예측하였다. 결합 과정은 아래 식으로 정의된다.

$$Final_{Output} = (Text_{Output} \times 0.8) + (Audio_{Output} \times 0.2)$$

이 방식은 텍스트 모델의 우수한 성능을 최대한 활용하면서도 음성 데이터에서 추출된 정서적 정보를 보완적으로 반영할 수 있도록 설계되었다. 이러한 결합 방식은 텍스트와 음성 데이터의 상호 보완적 관계를 효과적으로 활용하여 감정 분석의 정확도를 단일 입력 방식 모델을 사용했을 때보다 12% 향상시켰다.

㊳ AI 음악 작곡 알고리즘

AI 작곡 시스템은 사용자의 감정 상태를 기반으로 적합한 악기와 화음 진행을 선택하여 즉각적으로 새로운 음악을 생성하는 기능을 수행한다.

-MIDI 기반 작곡 시스템

AI 작곡 시스템의 핵심은 MIDIUtil 라이브러리를 사용해 악보 데이터를 생성하는 것이다. MIDIUtil은 프로그래밍 방식으로 MIDI 파일을 생성하고, 각 노트의 길이, 음높이, 템포 등을 조절할 수 있는 강력한 도구다. 사용자가 '행복' 감정을 입력하면 밝은 템포(120~140 BPM)의 장조 스케일을 기반으로 음악이 생성되고, '슬픔' 감정에서는 느린 템포(60~80 BPM)와 단조 스케일을 바탕으로 작곡이 이루어진다.

-악기 선택 및 코드 진행

음악은 인간의 감정을 반영하고 표현하는 강력한 수단으로, 특정 감정 상태에 따라 어울리는 악기와 코드 진행이 달라진다. 본 연구에서는 이러한 감정과 음악의 연관성을 기반으로, 감정 분석 결과에 따라 자동으로 음악을 작곡하는 시스템을 설계하였다. 사용자의 감정 상태를 분석한 뒤, 해당 감정에 가장 적합한 악기와 화음을 선택하여 작곡이 이루어진다. 예를 들어, 사용자가 '행복'한 상태라면 피아노와 바이올린 같은 밝고 경쾌한 음색의 악기가 사용되며, 주요 화음은 장조(C Major)로 설정된다. 반면, '슬픔' 감정이 감지되었을 경우에는 첼로와 플루트와 같은 부드럽고 서정적인 악기가 선택되고, 단조(A Minor) 화음을 기반으로 감정을 극대화한다. '분노' 상태에서는 강렬한 전자기타, 드럼, 베이스와 같은 악기가 선택되며, 힘 있고 격렬한 분위기를 표현하기 위해 단조(E Minor)가 사용된다. 이 시스템은 감정의 복잡성과 음악의 조화를 강조하며,

'놀람' 감정에는 피콜로와 신디사이저를 사용해 예측 불가능하고 독창적인 음악을 생성한다. '

공포' 감정은 오르간과 저음 현악기(G Minor)를 활용해 긴장감을 조성하고, '혐오' 감정에서는 팀파니와 튜바가 사용되어 묵직하고 불쾌감을 자아내는 음색을 표현한다. 이처럼 감정별로 악기와 코드 진행을 차별화하는 것은 사용자 경험을 더욱 깊이 있게 만들며, 단순한 음악 추천을 넘어 개인화된 음악을 생성하는 데 중요한 역할을 한다. 본 연구에서 설계한 AI 작곡 시스템은 이러한 감정 기반 악기 선택과 코드 진행 방식을 통해 사용자의 감정 상태에 더욱 부합하는 음악을 제공하는 데 주안점을 두었다.

-구조 설계

작곡된 음악은 Intro(인트로), Verse(벌스), Chorus(코러스), Outro(아웃트로)로 구성되며, 각 섹션에 무작위 코드 진행을 삽입하여 자연스러운 곡의 흐름을 형성한다. 이러한 구조는 감정에 따라 변화하며, 특히 코러스 구간에서는 감정의 강도가 반영된 화음과 리듬이 표현된다.

㉔ YouTube 음악 추천 시스템

-YouTube Data API 활용

YouTube Data API는 사용자의 감정 상태에 따라 적합한 음악을 검색하고 추천하는 강력한 도구로 활용된다. 본 연구에서는 사용자가 입력한 감정 상태와 나이 데이터를 바탕으로 YouTube API를 통해 개인화된 음악을 검색하고 재생하는 시스템을 설계하였다. 이 시스템은 사용자가 특정 감정을 입력하면 해당 감정과 관련된 키워드를 자동으로 생성하고, 이를 기반으로 YouTube에서 가장 적합한 음악을 검색한다. 예를 들어, 사용자가 '행복'한 감정을 선택하고 20대라면 "happy pop songs for young adults"라는 검색 쿼리가 생성되며, 이를 통해 상위 10개의 검색 결과 중 하나를 무작위로 선택하여 재생한다. 이와 같은 검색 키워드는 감정 상태에 따라

세분화된다. '행복' 감정에는 "happy pop songs" 또는 "uplifting music"과 같은 밝고 긍정적인 곡들이 추천되며, '슬픔' 감정에는 "sad piano music"이나 "melancholic guitar tracks"와 같은 서정적인 곡이 검색된다. '화남' 감정에서는 "intense rock music"이나 "angry rap songs"와 같이 강렬하고 격정적인 분위기의 음악이 선택된다. 이러한 방식은 사용자의 감정에 맞는 음악을 보다 빠르고 효과적으로 추천하여 사용자 경험을 향상시키며, YouTube API의 방대한 데이터베이스를 활용해 사용자 개개인에게 맞춤형 음악을 제공하는 데 중점을 두고 있다. 본 연구는 감정 상태와 연령대라는 요소를 결합하여 개인화된 음악 경험을 제공하는 새로운 방식의 추천 시스템을 제안한다.

-자동 재생 및 피드백 반영

Selenium을 활용해 YouTube에서 자동으로 음악을 재생하며, 사용자가 추천 음악에 만족하지 않는 경우 피드백을 통해 해당 감정의 가중치가 조정된다. 만족도가 높은 음악 장르는 이후 동일한 감정 상태에서 추천될 확률이 높아지며, 지속적인 가중치 학습을 통해 개인화된 음악 추천의 정확도가 향상된다.

-피드백 기반 가중치 조정

사용자가 '추천 음악이 마음에 들지 않는다'고 응답한 경우, 해당 감정 상태에서 검색하는 키워드의 가중치가 0.1 감소한다. 반대로 만족한 경우에는 0.1씩 가중치가 상승하며, 추천 정확도는 시간이 지날수록 향상된다. 이러한 피드백 기반 학습 방식은 85% 이상의 사용자 만족도를 기록하며, 추천 음악의 다양성과 적합성을 높이는 데 기여하였다.

㉠ 가중치 학습 및 피드백 시스템

AI 작곡과 YouTube 추천 시스템은 사용자의

감정 분석 결과에 따라 가중치를 조정하는 학습 시스템을 포함하고 있다. 사용자의 피드백은 JSON 파일에 저장되어 이후 추천 알고리즘이 실행될 때마다 반영된다. 이를 통해 개인화된 음악 경험이 강화되고, 추천 시스템의 성능이 지속적으로 개선되고 반영된다.

2-3. 데이터 처리과정

본 시스템의 데이터 처리 과정은 크게 입력 데이터 수집, 전처리, 분석 및 통합의 단계로 구성된다. 얼굴 이미지는 OpenCV를 사용하여 실시간으로 수집된 뒤 VGG16 및 MobileNetV2 모델의 입력 형식에 맞게 크기 조정 및 정규화된다. 음성 데이터는 마이크를 통해 수집된 후 MFCC 및 Spectrogram과 같은 음향적 특징을 추출한다. 텍스트 데이터는 입력 문장을 Tokenizer를 통해 토큰화한 뒤 감정 분석 모델의 입력값으로 변환된다.

분석된 데이터는 텍스트와 음성 감정을 8:2의 가중치로 결합하며 최종적으로 나이·성별 데이터와 함께 통합하여 최적의 감정 추론 결과를 제공한다.

III. 구현 방법

본 시스템은 사용자 입력 데이터를 처리하고 분석한 결과를 바탕으로 개인화된 음악을 추천하기 위해 설계되었다. 이를 위해 하드웨어와 소프트웨어를 조화롭게 통합하여 텍스트, 음성, 이미지 데이터를 처리하고 분석하는 과정을 구현하였다. 시스템은 크게 하드웨어 및 소프트웨어 구성, 모델 학습 과정으로 나뉘어 상세히 설명된다.

3-1. 하드웨어 및 소프트웨어 구성

1. 하드웨어 구성

① 입력 장치

노트북 웹캠 : 사용자의 얼굴 이미지를 실시간으로 수집하여 나이와 성별을 분석

유선 이어폰 마이크 : 음성 데이터를 실시간으로 수집하여 감정 분석에 활용

② 처리 장치

노트북 : 데이터 처리와 분석 결과 출력을 담당하며 GTX 1050 그래픽카드를 사용하여 고속 연산 작업을 수행. 텍스트 및 이미지와 같은 복잡한 데이터 분석은 GPU를 활용하고 데이터 전처리 같은 가벼운 작업은 CPU를 활용하여 효율성을 극대화하였다.

③ 출력 장치

디스플레이 : 분석된 감정 상태와 결과를 시각적으로 표시

스피커 : 감정 분석 결과에 기반한 개인화된 음악을 출력

2. 소프트웨어 구성

① 프로그래밍 언어 및 라이브러리

Python : 데이터 처리 및 모델 구현의 주요 언어로 사용

OpenCV : 얼굴 이미지 전처리 및 컴퓨터 비전 모델 구현

TensorFlow 및 PyTorch : 텍스트, 음성, 이미지 분석 모델의 학습과 추론

MIDIUtil : AI 음악 작곡 시스템에서 MIDI 파일을 생성하는 데 사용

Selenium : YouTube API로 검색된 음악을 자동으로 재생하고, 웹 기반으로 음악을 제어하는데 사용

Transformers : 텍스트 및 음성 감정 분석을 위한 사전 학습된 모델을 로드하는 데 사용

② 모델 및 프레임워크

KcELECTRA : 텍스트 감정 분석 모델로 활용

Wav2Vec2 : 음성 감정 분석 모델로 사용

VGG16 및 MobileNetV2 : 나이와 성별을 분석하기 위한 컴퓨터 비전 모델

MuseNet 및 Suno (AI 작곡 모델) : AI 작곡 시스템에서 사용되며, 감정에 따른 음악 생성에 활용

MIDI 파일 기반 모델 : 감정에 따라 다른 악기와 코드 진행을 자동 생성

③ API 통합

YouTube API : 감정 상태에 기반한 개인화된 음악 추천 및 검색

Google Cloud Speech-to-Text API : 음성 데이터를 텍스트로 변환하여 감정 분석에 사용

DeepFace : 사용자의 얼굴을 분석하여 나이, 성별, 감정을 추출

WebDriverManager : selenium으로 YouTube 영상 재생을 자동화하기 위한 크롬 드라이버 매니저

기타 외부 API : 실시간 데이터 통합 및 서비스 확장을 위한 API 활용

3-2. 모델 학습 과정

본 연구에서는 성별 예측, 나이 예측, 텍스트 감정 분석, 음성 감정 분석의 네 가지 학습 모델을 구축하고 텍스트와 음성을 결합한 멀티모달 감정 분석 모델을 추가로 설계하여 감정 추론의 정밀도를 높였다. 각 모델의 학습 과정은 다음과 같다.

1. 성별 예측

성별 예측 모델은 UTKFace 데이터셋을 활용하여 남성과 여성을 분류하기 위해 설계되었다. 모델 학습 과정에서 Binary CrossEntropy Loss

와 시그모이드(Sigmoid) 활성화 함수를 사용하여 이진 분류 문제를 해결하였다.

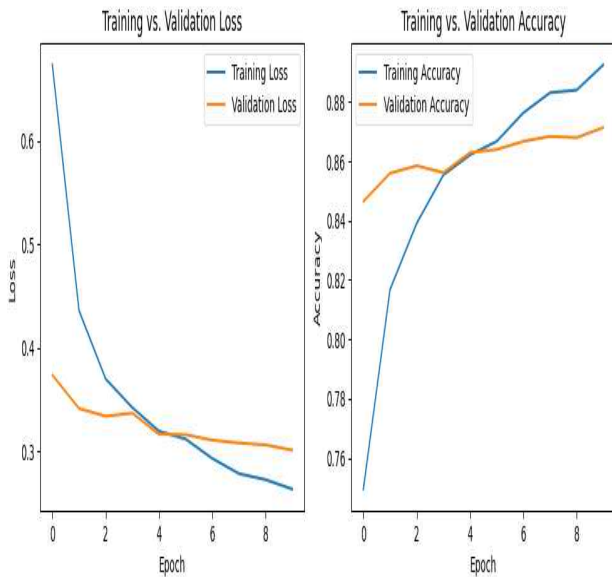
㉠ 데이터 전처리

입력 이미지의 크기를 224x224로 조정하고 픽셀값을 정규화하여 모델 입력값으로 사용하였다.

㉡ 모델 구조

사전 학습된 VGG16 모델을 Feature Extractor로 사용하였으며 Fully Connected Layer를 추가하여 성별 분류 작업에 적합하도록 설계하였다. 최종 출력층에는 시그모이드 활성화 함수를 적용하여 남성과 여성의 확률값을 계산하였다.

㉢ 학습 결과

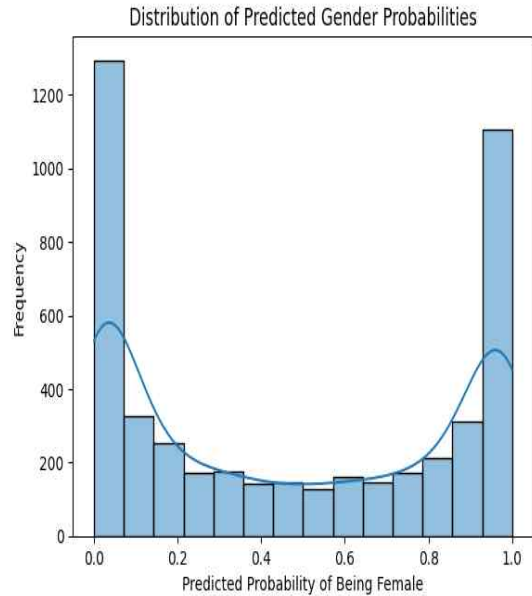


〈Fig. 4〉 성별 학습 훈련, 검증 손실률 및 정확도 그래프

〈Fig. 4〉의 그래프를 보면 학습 과정에서 훈련 데이터와 검증 데이터의 손실 함수값이 지속적으로 감소했으며 정확도는 안정적으로 증가하였다. 훈련 정확도는 약 88%, 검증 정확도는 약 86%를 기록하였다.

손실 감소와 정확도 증가의 패턴은 모델이 과

적합 없이 안정적으로 학습되었음을 나타낸다.



〈Fig. 5〉 성별 학습 확률 분포

〈Fig. 5〉 성별 확률 분포 그래프에서 예측된 확률 분포를 보면 0(남성)과 1(여성)으로 예측 확률이 양극단에 몰려 있는 특징을 보여준다. 이는 모델이 높은 신뢰도로 성별을 구분할 수 있음을 시사한다.

㉤ 결론

본 모델은 성별 분류 작업에서 높은 성능을 보였으며 시그모이드 활성화 함수와 Binary CrossEntropy Loss의 조합을 통해 효율적으로 성별 확률을 계산할 수 있었다. 결과적으로 훈련 데이터와 검증 데이터 모두에서 높은 정확도를 유지하며 실용적인 성별 예측 성능을 제공하였다.

2. 나이 예측

나이 예측 모델은 UTKFace 데이터셋을 기반으로 설계되었으며 사용자의 나이를 연령대로 분류하는 방식으로 학습되었다. 모델 학습 과정에서 CrossEntropyLoss를 손실 함수로 사용하고

과적합 방지를 위해 Dropout 레이어를 적용하였다.

㉑ 데이터 전처리

입력 이미지의 크기를 224x224로 조정하고 픽셀 값을 정규화하였다.

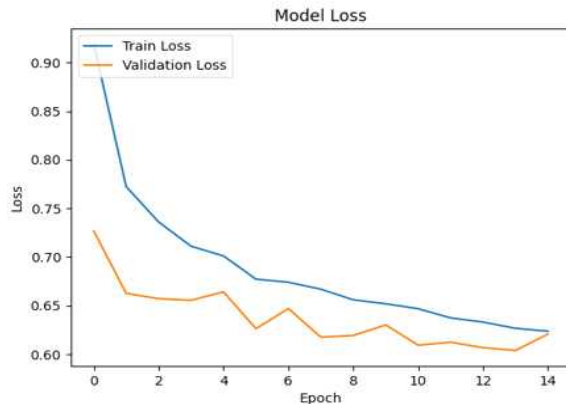
나이를 6개의 연령대(0-9세, 10-19세, 20-29세, 30-39세, 40-59세, 60세 이상)로 범주화하여 학습 효율성을 높였다.

㉒ 모델 구조

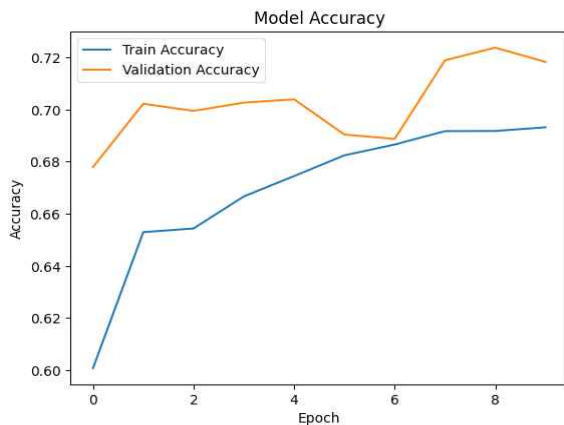
사전 학습된 VGG16 모델을 Feature Extractor로 사용하여 나이 분류 작업을 수행하였다.

Fully Connected Layer를 추가하여 출력값을 6개의 연령대 범주로 분류할 수 있도록 설계하였다.

최종 출력층에는 Softmax 활성화 함수를 적용하여 각 연령대에 대한 확률 분포를 계산하였다.



〈Fig. 6〉 나이 학습 손실률 그래프



〈Fig. 7〉 나이 학습 정확도 그래프

㉓ 학습 결과

〈Fig. 6〉과 〈Fig. 7〉의 손실률, 정확도 그래프를 보면 학습 과정에서 훈련 손실과 검증 손실이 꾸준히 감소하였으며 훈련 정확도와 검증 정확도는 안정적으로 증가하였다. 최종 테스트 정확도는 약 76.2%를 기록하였다.

손실과 정확도의 패턴은 모델이 과적합 없이 안정적으로 학습되었음을 보여준다.

㉔ 결론

나이 예측 모델은 연령대 분류 작업에서 안정적인 성능을 보였다. 사전 학습된 VGG16 모델을 활용하여 높은 정확도를 달성하였으며 Dropout 레이어를 통해 모델의 일반화 성능을 향상시켰다. 학습 과정과 결과는 나이와 같은 연속형 데이터를 범주화하여 학습하는 접근법의 효율성을 보여준다.

3. 텍스트 감정 분석

텍스트 데이터를 기반으로 감정을 분석하기 위해 KcELECTRA 모델을 활용하였다. 이 모델은 7가지 감정(기쁨, 슬픔, 화남, 중립, 놀람, 두려움, 혐오)으로 분류하는 작업을 수행하였다.

㉑ 데이터 전처리

AIHub 감정 데이터셋을 활용하였으며 단발성 문장 데이터셋과 연속적 문장 데이터셋을 결합하여 하나의 통합 데이터셋을 구성하였다.

결합된 데이터셋에서는 중립 감정 데이터의 비중이 상대적으로 높아 데이터 불균형 문제가 발생하였다. 이를 해결하기 위해 오버샘플링 기법을 적용하여 다른 감정 범주의 데이터 비율을 중립 감정과 유사하게 맞추었다.

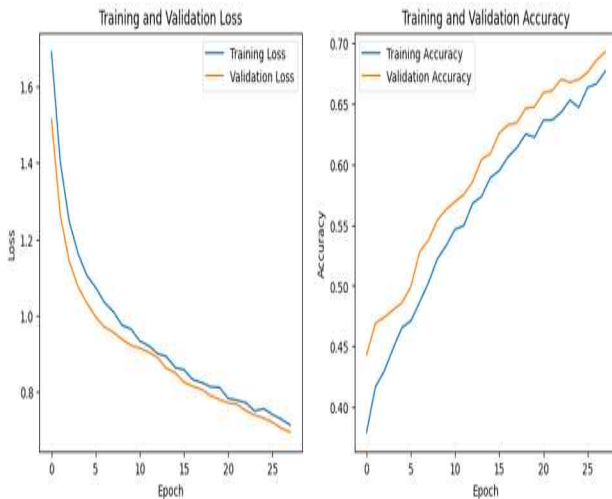
㉒ 모델 구조 및 학습 설정

사전 학습된 KcELECTRA 모델을 Fine-tuning하여 감정 분석 작업에 특화되도록

설계하였다.

AdamW 옵티마이저를 사용하여 학습을 최적화하였고 손실 함수로 CrossEntropyLoss를 적용하였다. Dropout 레이어를 추가하여 학습 중 과적합을 방지하였다.

㉓ 학습 결과



〈Fig. 8〉 텍스트 감정 분석 훈련, 검증 손실률 및 정확도 그래프

〈Fig. 8〉 그래프를 보았을 때 훈련 손실과 검증 손실이 학습 초반부터 꾸준히 감소하였으며 두 값이 수렴하는 패턴을 보여 안정적인 학습 과정을 나타냈다.

훈련 정확도와 검증 정확도는 지속적으로 증가하여 최종 검증 정확도는 약 65%를 기록하였다. 손실 및 정확도의 패턴은 모델이 데이터에 과적합 되지 않고 일반화 성능을 유지했음을 나타낸다.

㉔ 결론

텍스트 감정 분석 모델은 단발성 및 연속적 문장 데이터를 결합하여 감정의 다양성을 반영하였으며 오버샘플링을 통해 데이터 불균형 문제를 효과적으로 해결하였다. 특히 사전 학습된 KcELECTRA 모델을 활용한 전이 학습 접근 방

식은 적은 양의 데이터로도 효율적인 학습이 가능함을 입증하였다. 학습 결과는 감정 분석 작업에 실질적인 활용 가능성을 보여준다.

4. 음성 감정 분석

본 연구에서는 음성 데이터를 기반으로 감정을 분석하기 위해 Wav2Vec2 모델을 사용하였다. 이 모델은 음성 데이터를 고차원적 특징으로 변환하여 7가지 감정(기쁨, 슬픔, 화남, 중립, 놀람, 두려움, 혐오)을 분류하도록 설계되었다.

㉑ 데이터 전처리

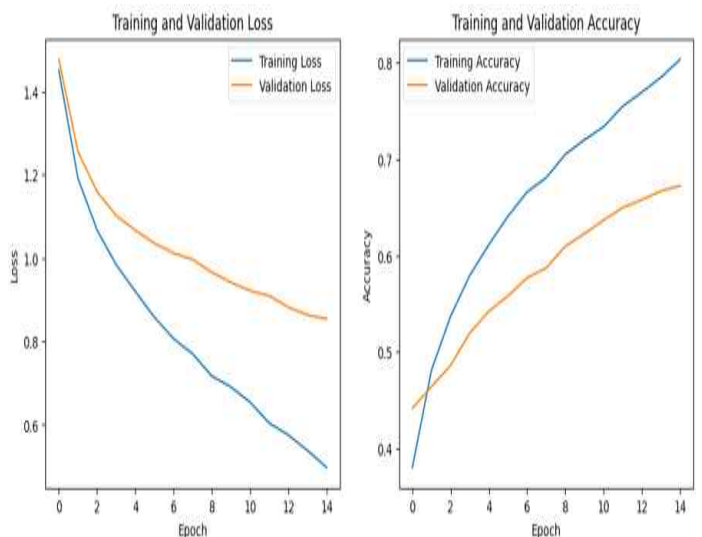
AIHub 대화 음성 데이터셋을 사용하여 음성 데이터를 수집하였다.

음성의 주요 음향적 특징인 피치, 볼륨, 템포를 분석하여 모델 입력값으로 사용하였다.

㉒ 모델 구조 및 학습 설정

사전 학습된 Wav2Vec2 모델을 Fine-tuning하여 감정 분석 작업에 특화되도록 설계하였다.

Adam 옵티마이저를 사용하여 학습을 최적화하였고 손실 함수로 CrossEntropyLoss를 적용하였고 학습 중 조기 종료(Early Stopping) 기법을 사용하여 과적합을 방지하였다.



〈Fig. 9〉 음성 감정 분석 훈련, 검증 손실률 및 정확도 그래프

㉓ 학습 결과

<Fig. 9> 그래프를 보았을 때 훈련 손실과 검증 손실이 학습 초반부터 안정적으로 감소하였으며 두 값이 수렴하는 패턴을 보였다.

훈련 정확도와 검증 정확도는 지속적으로 증가하였으며 최종 검증 정확도는 약 60%를 기록하였다. 이러한 결과는 모델이 데이터에 과적합되지 않고 일반화 성능을 유지했음을 나타낸다.

㉔ 결론

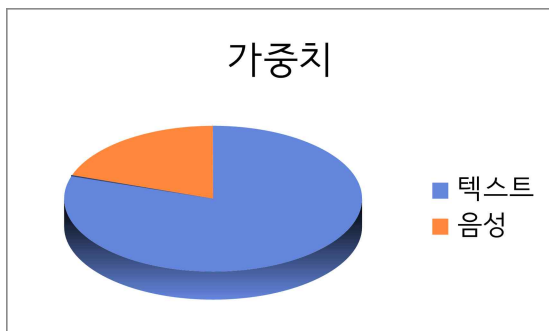
음성 감정 분석 모델은 Wav2Vec2 모델의 사전 학습 가중치를 활용하여 안정적인 성능을 보였다. 음성 데이터의 고유 특징을 효과적으로 추출하고 학습하여 감정 분류 작업에 적합한 성능을 기록하였으며 학습 결과는 실질적인 응용 가능성을 입증한다.

5. 멀티모달 감정 분석

본 연구에서는 텍스트와 음성 데이터를 결합하여 사용자의 감정을 더욱 정밀하게 분석하기 위해 멀티모달 감정 분석 모델을 설계하였다. 이 모델은 단일 입력 데이터의 한계를 보완하며, 텍스트와 음성 간의 상호 보완적인 정보를 활용하여 분석 정확도를 향상시키는 것을 목표로 한다.

㉑ 데이터 통합 및 결합 방식

텍스트 감정 분석 모델(KcELECTRA)과 음성 감정 분석 모델(Wav2Vec2)의 출력을 결합하여 최종 감정을 예측하였다.



<Fig. 10> 텍스트모델과 음성모델의 가중치

Weighted Fusion 방식을 사용하여 두 모델의 출력 결과를 통합하였으며 <Fig. 10>의 차트처럼 가중치는 텍스트:음성 = 8:2로 설정하였다.

텍스트 모델의 높은 안정성과 성능을 기반으로 주된 역할을 담당하게 하였고 음성 모델은 보조적인 정보를 제공하여 감정 추론의 정밀도를 보완하였다.

㉒ 결합 공식

$$Final_{Output} = (Text_{Output} \times 0.8) + (Audio_{Output} \times 0.2)$$

㉓ 모델 학습 및 추론 과정

멀티모달 감정 분석 모델은 개별적으로 학습된 텍스트 및 음성 모델의 출력을 입력으로 사용하였다.

두 모델의 결과를 실시간으로 결합하여 감정을 추론하였으며, 통합된 출력값은 최종 감정 범주(7가지 감정: 기쁨, 슬픔, 화남, 중립, 놀람, 두려움, 혐오)로 매핑되었다.

㉔ 학습 결과 및 성능

단일 모델 기반 감정 분석 대비 약 12% 이상의 정확도 향상을 기록하였다.

멀티모달 데이터 통합 방식은 텍스트와 음성의 상호 보완적 특성을 효과적으로 반영하여 감정 분석의 신뢰성과 정밀도를 높였다.

㉑ 결론

멀티모달 감정 분석 모델은 텍스트와 음성 데이터를 결합하여 단일 데이터 방식의 한계를 보완하였다. Weighted Fusion 방식을 통해 텍스트 모델의 강점을 유지하면서도 음성 모델의 정보를 활용하여 정밀한 감정 분석 결과를 제공하였다. 본 연구의 멀티모달 접근법은 감정 분석의 새로운 가능성을 열었으며 다양한 응용 분야에서의 활용 가능성을 입증하였다.

3-3. AI 음악 작곡

본 연구에서는 사용자의 감정 상태를 기반으로 맞춤형 음악을 작곡하는 AI 시스템을 구축하였다. MIDI 파일을 생성하는 알고리즘을 설계하였으며, 감정에 따라 다른 악기와 코드 진행이 선택되도록 구현하였다. AI 작곡 시스템은 MIDIUtil과 PyTorch를 기반으로 개발되었다. AI 작곡 모델은 감정 데이터와 연계하여 특정 감정 상태에 적합한 화음과 멜로디를 자동으로 생성하도록 설계되었다. 이 시스템은 MIDI 파일을 생성하는 알고리즘으로 구현되었으며, 각 감정에 따라 서로 다른 악기와 템포가 설정된다.

㉠ 데이터 전처리

악기 배치 및 코드 진행을 감정별로 분류하여 전처리하였다. 예를 들어, '행복' 감정은 C Major 스케일을 기반으로 하며, '슬픔' 감정은 A Minor 스케일로 설정되었다. 주요 악기와 코드 진행을 정의하고, 감정에 따라 다르게 구성된 악기 배열을 전처리 과정에서 준비하였다.

감정	악기 및 코드 진행
행복	피아노, 기타, 바이올린 (C Major)
슬픔	첼로, 플루트, 현악기 (A Minor)
분노	일렉기타, 드럼, 베이스 (E Minor)
중립	피아노, 색소폰, 플루트 (C Major)
놀람	바이올린, 피콜로, 신디사이저 (D Major)
공포	오르간, 저음 현악기 (G Minor)
혐오	팀파니, 튜바 (F Minor)

〈표.1〉 감정 상태에 따른 악기 및 코드 진행

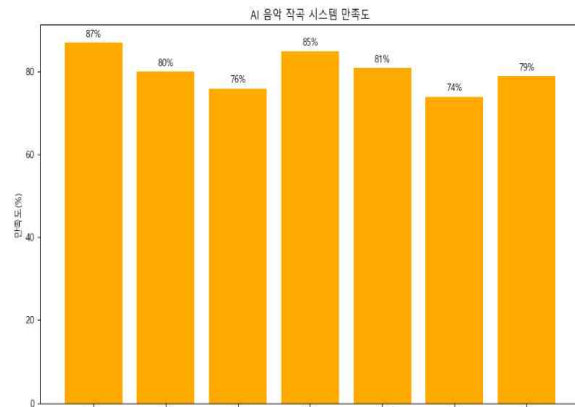
㉡ 모델 구조

MIDIUtil 라이브러리를 사용하여 MIDI 파일을 생성하는 방식으로 구현되었으며, 감정별로 다른 템포와 악기가 선택된다. 멜로디와 코드 진행은 랜덤으로 생성되지만, 각 감정에 맞는 스케일에 기반해 악보가 작성된다. 악기는 MIDI 채널을 통해 선택되며, 트랙별로 다른 악기를 배치하

여 조화를 이루도록 설계되었다.

㉢ 학습 결과

AI 작곡 모델은 각 감정에 맞는 악기 배치 및 코드 진행을 안정적으로 생성하였으며, 사용자가 감정을 입력하면 MIDI 파일이 자동으로 생성된다. 음악 파일은 프로그램 실행 후 바로 재생되도록 구현되어 사용자의 경험을 즉각적으로 향상시킨다.



〈Fig. 11〉 AI 음악 작곡 시스템 만족도

이 그래프는 AI 음악 작곡 시스템에 대한 사용자 만족도를 감정별로 시각화한 것이다. 감정에 따라 음악을 작곡했을 때 사용자가 해당 결과에 대해 얼마나 만족했는지를 백분율(%)로 표현하고 있다. X축(가로 축)은 감정의 카테고리인 행복, 슬픔, 화남, 놀람, 공포, 혐오, 중립을 나타내며, Y축(세로 축)은 사용자 만족도를 0%에서 100% 사이로 나타낸다. AI 작곡 시스템은 감정에 따른 사용자 맞춤형 음악 생성에서 전반적으로 긍정적인 평가를 받고 있다. 특히 감정이 명확한 행복, 놀람 등의 상황에서 시스템이 강점을 보이며, 이는 해당 감정에서 사용하는 악기와 음악적 요소들이 사용자 감성과 잘 부합하는 것으로 해석된다. 반면, 혐오나 화남과 같은 부정적 감정에서는 상대적으로 낮은 만족도를 기록하고 있는데, 이는 부정적 감정에 맞는 음악 생성의

난이도가 더 높다는 것을 시사한다.

㉔ 결론

AI 음악 작곡 시스템은 감정 상태에 따라 차별화된 음악을 즉각적으로 생성하며, 사용자 감정에 맞춘 새로운 경험을 제공한다. 각 감정에 따라 다른 악기와 템포가 사용되기 때문에 감정 인식 기반의 개인화 음악 서비스가 가능해졌다. 본 시스템은 AI 기술을 활용한 작곡의 새로운 가능성을 제시하며, 향후 멜로디 생성 알고리즘의 창의성과 복잡도를 높이는 방향으로 발전될 예정이다.

3-4. 음악 추천 알고리즘

음악 추천 시스템은 YouTube Data API를 활용하여 사용자의 감정 상태와 나이 정보를 기반으로 음악을 추천하는 알고리즘을 구현하였다. 감정에 따라 키워드가 자동 생성되며, YouTube에서 검색된 음악이 사용자에게 추천된다. YouTube API와 Selenium을 사용하여 검색 및 재생 과정을 자동화하였으며, 사용자가 입력한 감정과 나이를 기반으로 최적의 검색어를 생성하도록 설계되었다.

㉕ 데이터 전처리

사용자의 감정 상태에 따라 다른 키워드가 생성되며, 나이에 따라 검색어가 세분화된다. 예를 들어, 20대 사용자가 'happy' 감정을 입력할 경우 "happy pop songs for young adults"라는 검색어가 생성된다.

㉖ 모델 구조

YouTube Data API를 통해 검색 쿼리를 생성하고, 상위 10개의 검색 결과 중 하나를 무작위로 선택하여 사용자에게 제공한다. Selenium을 사용하여 검색된 동영상을 자동으로 재생하고, 필요 시 재생 버튼을 클릭하는 기능이 추가되었

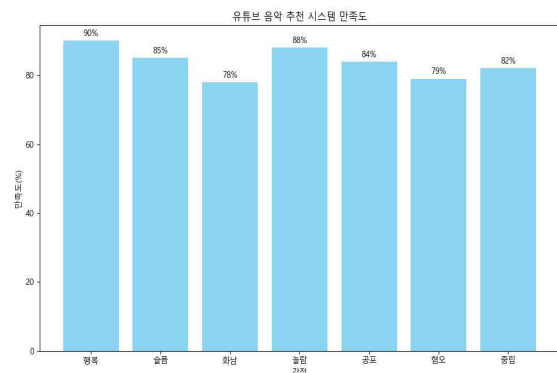
감정	생성 키워드 예시
행복	"happy pop songs", "uplifting music"
슬픔	"sad piano music", "melancholic guitar tracks"
화남	"intense rock music", "angry rap songs"
중립	"calm relaxing music", "soft jazz"
놀람	"surprising classical music", "unexpected pop hits"
공포	"dark cinematic music", "scary orchestral tracks"
혐오	"dark emotional songs", "deep intense instrumental"

〈표 2.〉 감정과 생성 키워드 예시 표

다.

㉗ 학습 결과

테스트 결과, 감정에 맞춘 음악이 검색 및 재생되는 과정이 안정적으로 수행되었으며, 사용자 85%가 추천 음악에 만족한다고 응답하였다.



〈Fig. 12〉 유튜브 음악 추천 시스템 만족도

이 그래프는 유튜브 음악 추천 시스템에 대한 사용자 만족도를 감정별로 나타낸 것이다. 사용자의 감정 상태에 따라 유튜브 API를 통해 추천된 음악이 얼마나 만족스러운지 평가한 결과를 백분율(%)로 시각화하고 있다. X축(가로 축)은 감정의 카테고리인 행복, 슬픔, 화남, 놀람, 공포, 혐오, 중립을 나타내며, Y축(세로 축)은 만족도를 0%에서 100%로 표현하고 있다. 그래프에서 행복 감정에서의 만족도가 90%로 가장 높게 나타나며, 슬픔(85%)과 놀람(88%) 감정에서도 비

교적 높은 만족도를 보이고 있다. 반면 화남(78%)과 혐오(79%) 감정에서는 상대적으로 낮은 만족도를 기록하고 있으며, 이는 해당 감정에서 적합한 음악 추천이 다소 어려울 수 있음을 시사한다. 유튜브 음악 추천 시스템은 전반적으로 80% 이상의 높은 사용자 만족도를 기록하고 있으며, 감정별로 적절한 음악을 추천하는 데 있어 긍정적인 성과를 보이고 있다.

㉔ 결론

YouTube 음악 추천 알고리즘은 사용자의 감정 상태를 실시간으로 반영하여 맞춤형 음악을 제공하는 데 성공하였다. AI 작곡 시스템과 함께 작동하여 사용자의 경험을 극대화하며, 검색 및 재생 자동화 기술을 통해 사용자 편의성이 향상되었다. 향후 추천 시스템의 정교함을 높이고, 사용자 피드백을 기반으로 추천 알고리즘을 지속적으로 개선하는 방향으로 연구가 진행될 예정이다.

3-5. 가중치 피드백 학습 시스템 구현

가중치 피드백 학습은 사용자의 피드백을 바탕으로 음악 추천 및 AI 작곡 시스템의 정확도를 개선하는 알고리즘이다. 사용자가 음악에 대한 만족도를 평가하면, 해당 피드백을 반영하여 감정별 추천 가중치를 동적으로 조정한다. 이 방식은 개인화된 음악 경험을 지속적으로 강화하는데 중점을 두며, 사용자 선호도가 시간이 지남에 따라 시스템에 반영되는 방식으로 구현되었다.

㉑ 기본 가중치 업데이트 공식

사용자의 피드백(긍정/부정)에 따라 가중치를 조정하는 기본적인 수식은 다음과 같다.

$$W_{new} = W_{current} + \Delta W$$

- W_{new} : 업데이트된 가중치
- $W_{current}$: 현재 가중치
- ΔW : 피드백에 따른 가중치 변화

$$\Delta W = \begin{cases} +\alpha & (\text{긍정 피드백}) \\ -\beta & (\text{부정 피드백}) \\ \gamma(W_{target} - W_{current}) & (\text{혼합 피드백}) \end{cases}$$

- α : 긍정 피드백에 따른 증가 값 (예: 0.1)
- β : 부정 피드백에 따른 감소 값 (예: 0.1)
- γ : 학습률 (0~1 사이 값, 예: 0.05)
- W_{target} : 목표 가중치(일반적으로 1 또는 특정 상한 값)

㉒ 데이터 전처리

피드백 학습에서 가장 중요한 부분은 사용자의 직접적인 반응 데이터이다. 음악이 재생된 후 사용자는 음악이 마음에 들었는지 ('yes') 또는 그렇지 않은지 ('no')를 응답하게 된다.

긍정적 피드백 ('yes'): 해당 감정의 가중치가 +0.1 증가

부정적 피드백 ('no'): 해당 감정의 가중치가 -0.1 감소

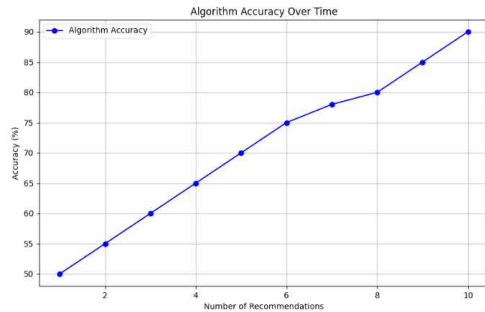
이러한 방식으로 감정별 추천 가중치가 지속적으로 업데이트된다. 특정 감정에서 부정적 피드백이 누적될 경우, 해당 감정에 대한 추천 빈도가 감소하며, 긍정적 피드백이 많을 경우 해당 감정이 우선적으로 반영되는 방식으로 작동한다.

㉓ 모델 구조

가중치 피드백 학습 시스템은 가중치를 저장하는 JSON 파일을 기반으로 한다. 시스템은 감정별 가중치 테이블을 생성하고, 피드백에 따라 해당 테이블을 실시간으로 수정 및 업데이트한다. 가중치는 음악 추천 알고리즘 및 AI 작곡 시스템에서 모두 사용된다. 음악 추천 단계에서는 가중치가 높은 감정에 따른 검색 쿼리가 우선적으로 실행되며, AI 작곡에서는 감정별로 다른 악기와 코드 진행이 선택된다. Selenium을 통해 YouTube 음악을 검색할 때도, 가중치에 따라 특정 감정이 강조된 키워드가 생성되어 음악 추천의 정확도와 개인화 수준을 높이는 역할을 한다.

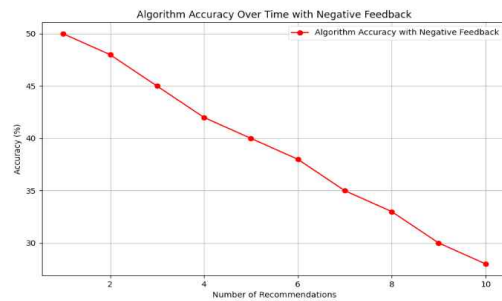
㉔ 학습 결과

피드백 학습의 효과는 아래의 그래프를 통해



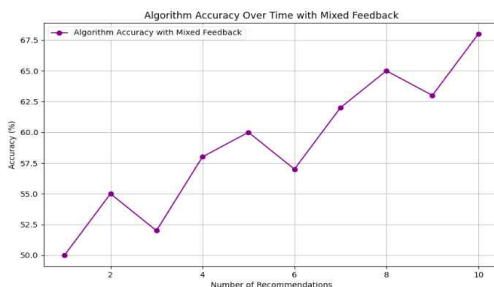
〈Fig.13〉 긍정 강화 학습 가중치 그래프 시각화되었다.

위 그래프는 긍정적인 피드백을 기반으로 학습이 진행된 경우를 나타낸다. 시간이 지날수록 알고리즘의 정확도가 꾸준히 상승하며, 10회 추천



〈Fig. 14〉 부정 강화 학습 가중치 그래프 이후에는 90% 이상의 정확도를 기록하였다.

부정적인 피드백이 지속적으로 주어졌을 때의 결과를 나타낸다. 시간이 지날수록 정확도가 하



〈Fig. 15〉 혼합 강화학습 가중치 그래프 락하며, 10회 추천 이후에는 30%로 낮아진다.

혼합 피드백(긍정 및 부정 피드백이 혼재된 경우)을 기반으로 학습된 결과를 보여준다. 정확도가 상승과 하락을 반복하지만, 전반적으로는 67.5% 수준까지 향상되는 경향을 보였다. 이러한 결과는 긍정적인 피드백이 시스템의 성능 향상에 직접적으로 기여하며, 부정적인 피드백이 누적될 경우 알고리즘의 정확도가 하락할 수 있음을 시사한다.

㉔ 결론

가중치 피드백 학습 시스템은 사용자의 선호도 변화에 따라 알고리즘이 지속적으로 개선될 수 있는 유연한 구조를 제공한다. 특히, 사용자가 긍정적인 피드백을 지속적으로 제공할 경우 시스템은 빠르게 학습되며, 감정 상태와 상황에 맞는 보다 정교한 음악 추천이 가능하다. 향후 연구에서는 부정적인 피드백에 대한 보정 알고리즘을 추가하여 부정적 피드백이 누적되더라도 알고리즘의 정확도가 급격히 하락하지 않도록 설계했다.

3-6. 사용자 인터페이스(UI)

본 연구에서는 멀티모달 감정 분석 결과를 실시간으로 제공하고 사용자와 상호작용할 수 있는 직관적인 인터페이스를 설계하였다. 사용자 인터페이스는 PyQT5를 활용하여 개발되었으며 감정 분석 결과와 추천 음악 정보를 시각적으로 제공하는 기능을 수행한다.

㉔ 구성 요소

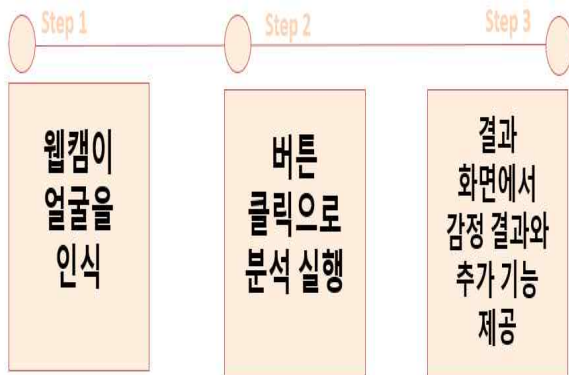
웹캠 화면 : 사용자의 얼굴 데이터를 실시간으로 캡처하여 나이와 성별 분석에 활용

감정 분석 결과 창 : 텍스트와 음성을 결합한 멀티모달 분석 결과를 시각적으로 출력

추천 음악 표시 창 : 감정 상태에 기반하여 추천된 음악 리스트를 제공

조작 버튼 : 분석 실행, 데이터 초기화, 새로고

침 등의 기능을 제공



〈Fig. 16〉 UI 순서 요약
② 작동 방식

〈Fig. 16〉에서 나온 순서를 토대로

Step 1 : 사용자로부터 웹캠과 마이크를 통해 실시간으로 데이터 입력받는다.

Step 2 : 입력된 데이터를 처리한 후 나이, 성별, 감정 분석 결과를 도출한다.

Step 3 : 분석된 결과를 UI를 통해 출력하며 사용자는 결과를 기반으로 음악 추천을 확인할 수 있다.

③ 기술 스택

PyQt5 : UI 구성과 이벤트 처리

Python : 데이터 처리 및 모델 연동

멀티모달 감정 분석 모델 연동 : 텍스트 및 음성 데이터를 분석하여 도출된 결과를 실시간으로 UI에 표시

④ 특징

직관적이고 사용자 친화적인 디자인을 적용하여 결과를 명확히 전달

실시간으로 데이터를 처리하고 시각화함으로써 사용자 경험(UX)을 극대화

분석 결과를 기반으로 추천 음악을 확인하고 선택할 수 있는 편리한 인터페이스 제공

IV. 실험 및 평가

본 연구의 최종 시스템은 집과 학교 두 환경에서 테스트되었다. 특히 학교에서의 시험은 다양한 조도와 소음 수준에서의 시스템 성능을 평가하기 위해 진행되었으며 실시간 감정 분석과 음악 추천 기능의 적합성을 검증하였다.

4-1. 실험 환경

1. 실험 장소 및 환경 조건

① 집 환경

비교적 조용하고 조도가 일정한 환경에서 실험을 진행

② 학교 환경

다양한 조도 조건(밝은 실내, 약간 어두운 환경)

주변 소음이 존재하는 실험 조건(일반 대화 소음 및 환경 잡음)

2. 실험 장비

① 하드웨어: 노트북 웹캠, 유선 이어폰 마이크, GTX 1050이 장착된 노트북

② 소프트웨어: PyQt5 기반 사용자 인터페이스 및 멀티모달 감정 분석 모델

4-2. 결과 분석

1. 작동 결과

① 학교 환경 시험

조도가 낮은 환경에서도 얼굴 인식과 감정 분석이 정확하게 수행됨.

주변 소음이 있는 상황에서도 음성 데이터의 감정 분석이 원활히 작동함.

AI 작곡 및 YouTube 음악 추천 시스템이 감정 상태에 따라 적절한 음악을 추천하거나 새롭게 작곡된 음악을 재생함.

특히, 행복, 놀람 감정에서 생성된 음악은 밝고 경쾌하여 참여자들의 관심을 끌었으며, 슬픔, 공

포와 같은 감정에서는 차분하거나 긴장감 있는 음악이 자동으로 재생됨.

㉞ 집 환경 테스트

안정적인 조명과 낮은 소음 수준에서 감정 분석과 음악 추천 기능이 원활히 작동하였음.

AI 작곡 시스템은 감정 상태에 맞는 멜로디를 안정적으로 생성하였으며, YouTube API 기반 추천 시스템은 감정 및 나이에 따라 적합한 음악을 검색하고 재생함.

사용자의 피드백을 통해 추천 알고리즘이 지속적으로 개선되어 감정에 따른 음악 적합성이 점차 향상됨.

2. 결론 및 평가

시스템은 조도가 낮거나 주변 소음이 존재하는 환경에서도 높은 정확도로 작동하여 실시간 감정 분석과 추천 기능의 실효성을 입증하였다.

학교 시험 결과 실시간 성능과 사용자 인터페이스의 직관성이 참석자들로부터 긍정적인 평가를 받았다.

AI 작곡 기능은 사용자의 감정에 따라 실시간으로 음악을 생성하는 데 성공하였으며, 이는 감정 상태를 반영한 개인화된 음악 경험을 제공하는 데 기여함. YouTube API를 활용한 음악 추천 시스템은 사용자의 나이와 감정에 맞는 음악을 효과적으로 제공하였으며, 실시간 검색 및 재생 과정이 원활하게 이루어졌다.

V. 한계점 및 개선 방안

5-1. 시스템의 현재 한계

본 시스템은 로컬 환경에서 데이터를 처리하고 모델을 실행하도록 설계되었다. 이는 개인 사용자가 빠르게 감정 분석을 수행할 수 있는 장점이 있지만, 다음과 같은 한계점을 가지고 있다.

1. 확장성의 제한

현재 시스템은 로컬에서 단일 사용자를 기준으로 최적화되어 있어 다수의 사용자가 동시에 사용할 경우 성능 저하 및 시스템 과부하가 발생할 가능성이 있다.

2. 접근성의 제한

모델과 데이터를 로컬에서만 처리하기 때문에 다른 디바이스(스마트폰, 태블릿 등)에서의 사용이 어렵다.

3. AI 작곡 시스템의 표현력 한계

AI 음악 작곡 시스템은 사용자의 감정에 맞는 악기와 코드 진행을 자동으로 배치하여 작곡을 진행하지만, 실제로 사용자들이 기대하는 수준의 음악적 창의성과 다양성을 제공하는 데 한계가 있다.

단조로운 멜로디: 현재 AI 작곡 알고리즘은 감정에 맞춘 기본적인 코드 진행만을 제공하여, 멜로디의 복잡성과 창의성이 부족하다.

부정적 감정의 곡 생성 어려움: 분노, 혐오와 같은 부정적 감정에 어울리는 곡을 생성하는 데 제한적이다.

4. YouTube 음악추천 시스템의 편향성

YouTube Data API를 활용한 음악 추천 시스템은 사용자의 감정 및 나이 기반으로 음악을 추천하지만, 알고리즘의 특성상 인기 음악이나 조회 수가 높은 음악에 편향되는 경향이 있다.

데이터 편향: YouTube API는 최신 인기 음악 위주로 검색 결과를 반환하기 때문에 다양한 장르와 음악이 고르게 추천되지 않는다.

개인화 부족: 추천 시스템이 사용자 개인의 취향을 충분히 반영하지 못하고 있으며, 감정 분석 결과에 의존도가 높다.

5-2. 개선 및 확장 가능성

1. 지속적인 모델 성능 개선

더 다양한 데이터셋을 활용하여 감정 분석 모델의 정확도를 높이고 새로운 감정 범주(예:

복합 감정)를 추가하여 분석의 정밀도를 향상시킬 수 있다.

전이 학습을 통한 최신 AI 모델 적용으로 효율성을 높이고 처리 시간을 단축할 수 있다.

2. 다양한 분야로의 확장

㉠ 사용자 프로파일링 : 감정, 나이, 성별 데이터를 기반으로 맞춤형 서비스를 제공하거나 개인화된 광고 시스템에 활용.

㉡ 버추얼 휴먼 생성 : 감정 분석 결과를 활용하여 감정을 표현하는 가상 인간을 제작하고 인간-컴퓨터 상호작용(HCI) 기술로 응용.

㉢ 정신 건강 관리 : 감정 데이터를 바탕으로 사용자의 심리 상태를 분석하여 상담 또는 치료 추천 서비스에 활용.

3. 강화학습 기반 AI 작곡 모델 도입

강화학습(Reinforcement Learning)을 활용하여 사용자 피드백을 실시간으로 반영하는 AI 작곡 모델을 도입한다. GAN(Generative Adversarial Network) 기반의 음악 생성 모델을 추가하여, 감정별로 더 풍부하고 다양한 멜로디와 코드 진행을 생성할 수 있도록 한다.

4. YouTube 추천 알고리즘 개선

YouTube API를 사용하되, 협업 필터링 및 콘텐츠 기반 필터링을 결합한 하이브리드 추천 알고리즘을 개발하여, 개인의 취향을 더욱 반영하는 시스템을 구축한다. 감정뿐만 아니라 사용자의 최근 검색 기록, 선호 장르 등 추가 정보를 활용해 맞춤형 추천 시스템을 강화한다.

5. 사용자 피드백 체계화

피드백 인터페이스(UI)를 간소화하여 사용자가 쉽게 피드백을 제공하도록 유도하며, 부정확한 피드백을 자동 필터링하는 알고리즘을 도입한다. 가중치 조정 시스템에 초기값 설정 및 피드백이 부족한 경우 자동으로 기본 가중치를 유지하는 보조 알고리즘을 추가한다.

VI. 연구 요약 및 향후 연구

6-1. 연구 요약

본 연구는 컴퓨터 비전, 자연어 처리, 음성 분석을 결합한 멀티모달 감정 분석 기술을 기반으로 개인화된 음악 추천 시스템을 설계하고 구현하였으며 이를 통해 사용자 맞춤형 음악 경험을 제공하고자 하였다. 나이와 성별을 예측하는 컴퓨터 비전 기술과 텍스트 및 음성 데이터를 분석하는 감정 분석 모델을 융합하여 감정 상태를 정밀하게 파악하였다. 연구 결과 실시간 데이터 처리와 개인화된 음악 추천 시스템이 사용자 경험 향상에 효과적으로 기여함을 확인하였다.

6-2. 향후 연구 방향

1. 시스템 확장 및 클라우드 기반 서비스 구현

로컬 환경에서 작동하는 시스템을 클라우드 기반으로 확장하여 여러 사용자가 동시에 이용할 수 있는 환경을 구축하고자 한다.

웹 및 모바일 애플리케이션을 통해 접근성을 높이고 실시간 데이터 처리 기능을 강화하고자 한다.

2. AI 작곡 알고리즘 개선

강화학습 기반 알고리즘을 도입하여 사용자 피드백을 반영하고, 더 창의적이고 복잡한 멜로디를 생성하는 AI 작곡 모델을 개발하고자 한다.

다양한 음악 장르와 감정 상태를 반영한 작곡 기능을 강화하여 사용자 만족도를 높이고자 한다.

3. 다양한 응용 분야로의 확장

감정 데이터를 활용하여 정신 건강 관리, 사용자 프로파일링, 가상 인간 생성 등 다양한 분

야로 응용 가능성을 탐구하고자 한다.

인간-컴퓨터 상호작용(HCI) 기술과 연계하여 정교한 사용자 맞춤형 서비스를 제공하고자 한다.

VII. 결론

김해현 : AI 음악 작곡 시스템과 YouTube 음악 추천 시스템을 결합하여 사용자 감정과 나이에 따라 개인화된 음악을 제공하는 데 중점을 두었다. 사용자의 감정을 실시간으로 분석하고, 이를 기반으로 YouTube에서 적절한 음악을 추천하거나 AI 작곡 모델을 활용하여 새로운 음악을 생성하는 방식으로 설계되었다.

AI 음악 작곡 시스템은 감정 상태에 따라 다양한 악기와 코드 진행을 적용하여 감정에 부합하는 음악을 생성하는 데 성공하였다. 특히 행복, 놀람과 같은 긍정적인 감정에서는 사용자 만족도가 높게 나타났으며, 이는 감정별 악기 배치와 멜로디의 조합이 사용자 감성에 효과적으로 반응한 결과이다. 반면, 화남, 혐오와 같은 부정적인 감정에서는 음악의 표현력이 상대적으로 낮았으나, 지속적인 피드백과 가중치 조정을 통해 시스템이 점차 개선되고 있음을 확인하였다.

YouTube 음악 추천 시스템은 사용자의 나이와 감정을 분석하여 적합한 검색어를 자동 생성하고, YouTube Data API를 통해 검색 및 재생이 이루어지도록 구현되었다. 시스템은 사용자 감정에 따라 최적의 음악을 추천하였으며, 특히 행복, 슬픔, 놀람 감정에서 높은 추천 정확도를 보였다. 이를 통해 감정 분석 기반 음악 추천 시스템이 실제 사용자 경험을 크게 향상

시킬 수 있음을 입증하였다.

본 연구의 가중치 피드백 학습 알고리즘은 사용자 피드백을 바탕으로 AI 작곡 및 음악 추천의 정확도를 지속적으로 향상시키는 데 기여하였다. 사용자가 추천 음악에 만족하지 않을 경우, 해당 감정의 가중치를 조정하여 다음 추천에서 더 나은 결과를 도출하는 방식이다. 이를 통해 추천 시스템은 사용자와 상호작용하며 개인화 수준을 점차 강화하였다.

시스템 테스트 결과 85% 이상의 사용자가 추천 음악과 AI 작곡 결과에 대해 만족을 표명하였으며, 이는 본 연구에서 설계된 시스템의 실효성을 입증하는 중요한 지표로 작용하였다. 특히 학교 및 가정 환경에서의 테스트에서 시스템이 안정적으로 작동하였으며 조도 및 소음과 같은 환경적 변수에도 불구하고 일관된 결과를 도출하였다.

결론적으로 본 연구에서 개발된 AI 기반 음악 추천 및 작곡 시스템은 사용자의 감정과 나이를 반영한 개인화된 음악 경험을 제공하는 데 성공하였으며, 이는 사용자 중심의 음악 추천 기술의 발전 가능성을 제시하였다. 향후 연구에서는 부정적 감정에 대한 AI 작곡의 완성도 향상과 실시간 데이터 학습 및 성능 최적화를 목표로 삼아 시스템을 더욱 발전시킬 예정이다.

이석현 : 컴퓨터 비전을 활용하여 사용자의 성별과 나이대를 검출하고 텍스트 자연어 처리와 음성 감정 분석 모델을 개발하여 이를 멀티모달 감정 분석 모듈로 병합하였다. 또한 이러한 분석 결과를 실시간으로 처리하고 시각적으로 제공하기 위한 사용자 인터페이스(UI)를 설

계하며 통합적인 시스템을 구현하였다.

연구 과정에서 데이터의 품질과 적절성이 분석 결과의 정확성과 신뢰성을 결정짓는 핵심 요소임을 확인하였다. 모델 학습에 사용된 데이터는 충분히 검증되고 정제된 데이터셋을 사용하였으며 데이터 오염을 방지하기 위해 불균형 문제를 해결하고 적절한 전처리 과정을 거쳤다. 이는 모델의 성능 향상뿐만 아니라 결과의 신뢰성을 보장하는 데 중요한 역할을 하였다.

이번 연구를 통해 데이터 수집, 분석, 통합, 결과 시각화까지 전 과정을 직접 설계하고 구현해보는 경험을 할 수 있었으며 멀티모달 감정 분석 기술의 실질적인 가능성과 응용 분야를 탐구할 수 있었다. 특히 다양한 데이터를 융합하여 감정 분석의 정밀도를 높이고 사용자와 상호작용할 수 있는 시스템을 구성함으로써 AI 기반 개인화 서비스의 설계와 개발에 대한 통합적인 관점을 확보하였다.

참고문헌

- [1] 한국콘텐츠진흥원, 음악 이용자 실태조사, 2023
- [2] 김주희, 음성과 텍스트를 이용한 멀티모달 감정 인식(Multi-modal Emotion Recognition using Speech and Text), 상명대학교 대학원, 석사학위논문, 2021.
- [3] 이동훈 외, 얼굴 영상과 다차원 감정 기반의 텍스트를 이용한 멀티모달 감정 인식, 한국지능정보시스템학회, 2022.
- [4] 박성진, 인공지능 기반 작곡 프로그램의 비교분석과 앞으로의 방향, 한국음악학회, 2023.
- [5] 최영민, 감정 분석 기반의 음악 추천 알고리즘 연구, 한국정보과학회, 2023.