

운전자 감성인지 기반 맞춤형 음악 생성 AI 인포테인먼트

김도연, 손봉기*, 이재호
덕성여자대학교, *서원대학교

ehdus2022@duksung.ac.kr, *bksohn@seowon.ac.kr, izeho@duksung.ac.kr *

AI Infotainment for Custom Music Generation Based on Driver's Emotional Recognition

Kim Do Yeon, Bong-Ki Son*, Jaeho Lee
Duksung Woman's Univ., *Seowon Univ.

요약

본 논문은 운전 중 무료함을 해소하기 위해 운전자의 감성 상태와 환경적 요인을 인지하여 운전자 맞춤형 음악을 생성하는 AI 인포테인먼트 구현 방법에 대해 연구하였다. 본 연구에서 제안하는 음악 생성 AI 인포테인먼트의 음악은 운전자의 표정 및 환경 정보를 활용하여 생성한 멜로디와 운전자의 User Intent 를 반영하여 생성한 편곡으로 이루어진다. 이때, User Intent 는 운전자의 표정을 Input 으로 하여 음악 Playtime 을 오래 지속하는 방향으로 강화학습 한다. 이를 위해 세 가지 주요 모델, 즉 운전자 감성 인지 기반 멜로디 생성 모델, 사용자 반응형 편곡 모델, 그리고 음악 Playtime 지속을 위한 강화학습 모델을 제안한다.

I. 서론

차량 내 인포테인먼트(IVI) 시스템의 성장은 운전자에게 다양한 혜택과 가치를 제공할 수 있기에[1], 최근 차량 내 인포테인먼트 시스템과 함께 운전자의 운전 경험을 향상하기 위한 다양한 기술이 연구되고 있다.

특히, 운전 중 무료함을 달래기 위한 맞춤형 음악 서비스는 운전자의 운전 경험을 향상할 수 있는 요소이다. 본 논문은 음악 감상 요구가 아닌, 운전의 무료함을 달래기 위한 단순 음악 청취 요구를 가진 운전자의 맞춤형 AI 음악 스트리밍 서비스에 집중한다.

운전자는 스트리밍 서비스 이용료나 차량 자체에 부과되는 통신료 등의 금전적 지출을 지양할 수 있으며, 휴대전화 연동 및 유료 서비스에 대한 의존성은 음악 서비스 수요를 제한할 수 있다. 또한, 무작위로 재생되는 AI 음악을 도입할 경우, 운전자의 취향과 상황에 따른 감성 불일치에 의해 청취 경험의 향상이 제한적일 수 있다.

따라서, 본 연구에서 제안하는 음악 생성 AI 인포테인먼트는 다음과 같은 효과를 기대한다. 차량 내 stand-alone 및 무료 서비스로, 기존 맞춤형 음악 서비스에서 필수적으로 요구되는 휴대전화 연동, 유료 음악 스트리밍 서비스, 그리고 저작권 문제를 겪지 않고, 비용 절감의 효과를 기대한다.

또한, 운전자의 표정 및 환경 정보를 활용하여 멜로디를 생성하며, 운전자 감성 맞춤을 위해 강화학습 체계를 가진 편곡을 생성함으로써 청취 경험 향상을 기대한다. 나아가, 정량 평가가 어려운 음악 분야의 강화학습 체계 구축을 위해, 사용자 관점의 보상 체계를 확립하여 AI 음악 생성 모델 강화학습의 선행 시스템을 제안하고자 한다.

II. 본론

2.1 차량용 인포테인먼트 시스템

인포테인먼트(Infotainment)란 정보(Information)와 오락(Entertainment)의 합성어다. 즉 차량용 인포테인먼트 시스템(In-Vehicle Infotainment)은 차량 내의 운전자나 동승자에게 여러 가지 정보나 오락적인 기능을 모두 제공할 수 있는 시스템을 의미한다[2]. 자율주행차와 커넥티드 카 시장이 성장함에 따라, 차량용 인포테인먼트 시스템은 운전자에게 자동차 안에서 새로운 경험을 누릴 수 있게 하는 핵심적인 요인으로 주목된다[1].

2.2 음악 생성 AI 인포테인먼트 개요

본 연구에서 제안하는 음악 생성 AI 인포테인먼트는 그림 1 과 같이 세 가지의 단계로 이루어진다. 멜로디 생성, 편곡 생성, User Intent 를 이용한 편곡의 분위기 변경이다.

첫째, 운전자의 감성을 반영하는 Color 를 기반으로 멜로디 생성 모델이다. 이 모델은 운전자의 현재 환경정보와 표정, 그리고 다양한 외부 요소를 분석하여 운전자의 감성을 인지한다. 둘째, 편곡 생성 모델은 생성된 멜로디를 기반으로 하며, 운전자의 분위기에 맞는 편곡을 제공할 수 있도록 User Intent 를 반영하여 강화 학습의 체계를 갖춰 편곡을 생성한다. 셋째, 강화 학습 모델은 편곡 생성 모델에 대한 강화 학습 체계를 갖추도록 운전자의 state action 을 결정한다. 이때 state action 은 사용자의 얼굴 이미지와 음악의 Playtime 을 통해 강화 학습하며 Playtime 이 지속되도록 결정한다. state action 은 편곡 생성 모델의 User Intent 로써 활용한다.

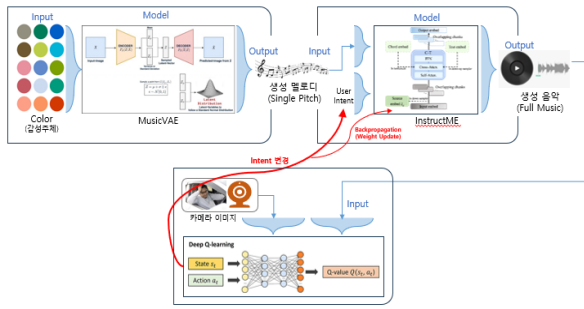


그림 1. 음악 생성 AI 인포테인먼트 구조

2.3 멜로디 생성 모델

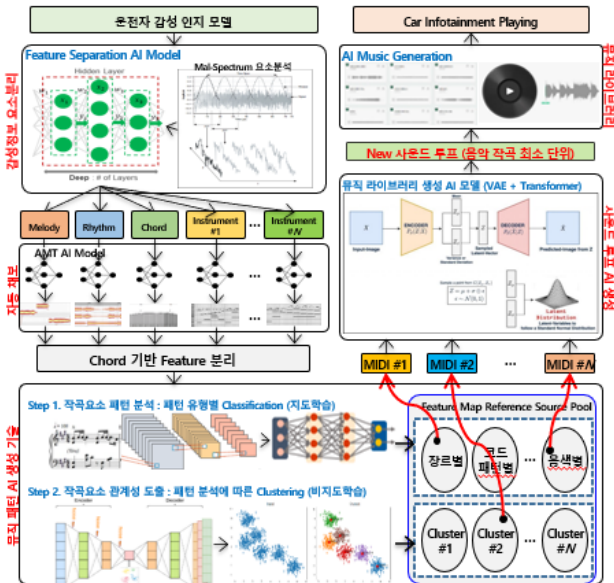


그림 2. 멜로디 생성 모델 구조

멜로디를 생성하는 MusicVAE 를 기반으로, 장기적인 음악적 구조를 이해하고 사용자가 지정한 조건(감정)을 반영하는 색상(Color)에 기반하여 개인화된 음악을 생성하는 모델이다. 또한, Actor-Critic Training 을 사용하여 학습 과정을 최적화하고, 음악에 대한 이해도가 높은 MusicVAE 의 latent vector 를 재학습한다.

먼저 Mel-spectrogram 데이터를 사용하여 음악 데이터 세트를 처리한다. 데이터 세트는 2D CNN 인 Mood Classifier 를 통해 음악의 감정에 따라 분류하고, 각 분류에 따른 Color 를 정형화(Color, Brightness, Saturation 등)한다. 각 감정에 따른 음악의 특징(pitch, density)을 추출한 후, 이를 Loss Function 으로 정의한 뒤, MusicVAE 를 통해 latent vector 를 사전 학습하고, Latent space 에서 Actor-Critic 방법으로 선택된 음악의 특징에 맞는 latent vector 를 추출해 재학습한다. 재학습한 latent vector 를 MusicVAE 의 decoder 에 입력하면 맞춤형 멜로디가 생성된다.

해당 모델은 사용자의 상황과 선호에 따라 다양한 조건을 설정하여 맞춤형 멜로디를 생성할 수 있고, 조건부 생성을 통해 더 독창적이고 개인화된 음악 청취 경험을 제공한다.

2.4 편곡 생성 모델

편곡 생성 모델은 편곡 모델 InstructME 를 기반으로, 2.3 절에서 기술한 멜로디 생성 모델이 도출한 멜로디에 적합한 편곡을 생성한다. 운전자 감성에 맞는 음악을 만드는 것은 멜로디보다 편곡에 의한 분위기 변환 역할의

비중이 높기 때문에, 생성된 멜로디를 토대로 운전자 감성 맞춤형 분위기를 갖기 위한 편곡 생성 모델이 필요하다.

편곡은 멜로디와 User Intent 를 통해 생성한다. User Intent 는 운전자의 표정을 기반으로, 음악의 Playtime 을 지속하도록 생성되는 action 이다. 기존 InstructME 모델은 condition 정보로써 chord progression 과 사용자 입력의 text description 을 요구하지만, 운전자 취향 맞춤형 최적화를 위해 text description 대신, User Intent 강화학습 모델의 Action 값을 condition 으로 사용한다. 또한, User Intent 강화학습 모델의 Loss 값을 변형하여 편곡 생성 모델의 Loss 와 Backpropagation 절차에 반영한다. 이는 편곡의 분위기를 운전자에게 더욱 적합하게 하는 강화학습 체계를 만들어준다.

2.5 User Intent 도출을 위한 강화학습 모델

User Intent 도출을 위한 강화학습 모델은 생성 음악의 Playtime 이 증가하는 것을 목적으로 state action 을 결정하는 face image 기반의 강화학습 모델이다.

운전자 감성 인지는 차량 환경에서 제한적일 수 있으며, 감성에 적합하도록 편곡의 분위기를 결정하는 방법도 운전자에 따라 다양할 수 있다. 그렇기에 표정 및 환경 정보만으론 어려울 수 있는 운전자의 감성인지를 state action 을 활용한 User Intent 를 통해 실현한다.

이에 본 모델은 음악의 Playtime 이 운전자가 듣고 싶은 음악을 결정할 때 중요 요소로 판단하고, 음악의 Playtime 을 지속할 수 있도록 강화 학습하며 state action 을 결정한다. state action 은 멜로디와 편곡 생성이 완료된 Full Music 을 대상으로, 사용자 취향 및 의도가 포함된 음악 Playtime 의 생존 비율을 극대화하기 위한 형태를 취한다.

User Intent 도출을 위한 강화학습 모델은 DQN(Deep Q-learning Network) 기반의 강화학습 모델로, Reward 는 생성음악 Playtime 생존시간 비율(played time / full music time), Input 값은 시간 종속성을 가진 운전자의 얼굴 이미지 4~6 장으로 하여, 음악 편곡모델(연구 주제 2)의 User Intent(Condition)가 되는 Action 을 결정한다. 나아가 PER(Prioritized Experience Replay) 알고리즘을 적용하여 학습에 영향을 준 주요 요소를 반복 학습하여 강화 학습의 효과를 높인다.

이때, DQN Loss Function 재설계가 필요하고, 편곡 생성 모델(연구 주제 2)의 원 모델인 MusicVAE 의 Loss 및 Backpropagation 또한 변형이 필요하다.

이 모델은 운전자 각 개인의 취향과 표정 변화에 맞도록 편곡 생성 모델을 강화 학습하여 맞춤형 음악 서비스를 제공하는 것을 목표로 한다.

III. 결론

본 논문은 운전 중 무료함을 해소하기 위해 음악 청취를 요구하는 운전자들을 위한 음악 생성 AI 인포테인먼트를 제안한다. 본 논문에서 제안하는 음악 생성 AI 인포테인먼트를 통해, 운전자는 차량 자체의 stand alone AI 음악 스트리밍으로, 운전 중 선곡에 주의를 두거나 음악 재생에 부과되는 비용 부담을 가지지 않고 단순 음악 청취에 대한 요구를 해결한다. 또한, 여러 운전자가 차량을 사용할 경우에도, 각 운전자의 감성을 인지하여 개인 맞춤형의 음악 서비스를 제공한다.

나아가 본 논문에서 제안하는 음악 생성 AI 인포테인먼트는 자율주행 서비스 활성화와 함께 더 큰 가치 창출을 기대할 수 있다. 또한, AI 음악 스트리밍에 국한하지 않고 다양한 음악 서비스로 발전할 수 있으며, 운전자 개개인의 감성 맞춤을 위해 DQN 모델을 적용한 강화 학습

체계를 구축함으로써, 사용자 관점의 보상 체계를 확립한
음악 생성 AI 인포테인먼트를 제안한다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보 통신부)의 재원으로
한국연구재단의 지원을 받아 수행된 연구임(과제번호-
2022R1A2C1009951).

참 고 문 헌

- [1] 김정환, 김민성. "차량용 인포테인먼트 시스템의 주요 속성에
대한 이용자 선호 분석." 디지털융복합연구 20.5 (2022):
617-626.
- [2] H. K. Shin (2019). Autonomous vehicle era, direction of
infotainment development. National IT Industry
Promotion Agency Issue Report, 2019-33, 1-13.