

사용자의 음악 선호요인 분석을 통한 개인 맞춤형 음악추천모델 연구

정지윤 · 김명준*

이화여자대학교 융합콘텐츠학과

A Study on Personalized Music Recommendation Model through Analysis on Users' Music Preference Factors

Ji-Yun Chung · Myoung-Jun Kim*

Department of Content Convergence, Ewha Womans University, Seoul 03760, Korea

요약

최근 인공지능 스피커의 대중화와 음악 스트리밍 서비스의 확장은 개인화 음악추천시스템의 중요성을 높였다. 본 연구는 음악 데이터 및 청취자의 선호 음악요인을 분석하여, 개인화된 콘텐츠기반 음악추천모델을 제안한다. 현재 사용되고 있는 음악추천시스템들은 청취 기록 같은 사용자의 데이터를 근거로 음악을 추천하고 있다. 그러나 이러한 방식은 기존 청취자의 기록이 없거나 새로운 음악의 경우에 선호도를 예측하기 어려운 한계가 있었다. 이 문제를 해결하기 위해 본 연구는 음악적 분석데이터를 활용하여 음악요소에 따라 개인별 음악 선인지여부와 선호여부를 알아보고, 사용자의 선호경향을 파악하여 선형 회귀분석을 통한 개인화된 음악추천모델을 제안한다. 그 결과, 본 연구의 추천모델은 음악요소 데이터와 개인의 선호여부만을 가지고도 음악 선호요인을 도출할 수 있었으며, 84.4% 높은 예측정확도를 보였다. 향후 본 연구는 음악 추천분야 뿐만 아니라 다양한 개인화 추천시스템에 적용될 수 있는 범용가능성을 확인하였다.

[Abstract]

Recently, popularization of AI speaker and expansion of Music streaming service have increased the importance of personalized Music recommendation system. In this study, we analyze music data and preferred music factors of listeners, and propose a personalized content based Music recommendation model. Music recommendation system currently in use recommends music based on user data similar to the user's listening record. However, this method has limitations in predicting the preference when there is no record of the existing listener or new music. To solve this problem, this study investigates the pre-recognition and preference of individual music according to musical factors using musical analysis data, and proposes personalized Music recommendation system through linear regression analysis by understanding the user's preference trends. As a result, the Music recommendation model derived the preference factor of music by music analysis data and personal preference alone, and showed the high prediction accuracy of 84.4%. In the future, this study confirms the applicability that can be applied to various personalized recommendation systems.

색인어 : 음악추천, 개인화 추천, 콘텐츠 기반, 음원추천, 머신러닝

Key word : Music Recommendation Model, Personalized Recommendation, Contents Based, Music Resource Recommendation, Machine Learning

<http://dx.doi.org/10.9728/dcs.2018.19.11.2041>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 10 October 2018; Revised 20 October 2018

Accepted 20 November 2018

*Corresponding Author; Myoung-Jun Kim

Tel: +82-2-3277-3347

E-mail: mjkim@ewha.ac.kr

1. 서론

최근 국내 대중음악은 K-pop의 인기와 더불어 디지털 음원 매출의 증가로 음원산업이 성장하고 있다. 음악 추천서비스가 중요한 이유는 음원에 대한 소비가 증가할수록 음원 추천을 통해 음악을 감상하는 경우가 많기 때문이다. 그러나 국내 5대 음원사이트 추천 곡의 신뢰여부에 대한 이용자 설문결과는 유용성과 신뢰성이 낮았으며, 음원차트의 추천음악을 신뢰하지 않는 것으로 나타났다[1]. 추천음악의 신뢰성이 낮은 원인은 음반사와 유통사의 일방적인 추천으로 취취자의 성향보다는 마케팅에 의존한 음원추천서비스가 이루어졌기 때문이다. 그리고 상용화된 음원추천시스템은 협업기반 필터링 시스템을 이용하기 때문에 개인의 성향보다는 다수의 선호도를 바탕으로 선호하는 음원을 예측하고 추천한다. 본 연구는 실험을 통해 개인의 음악적 선호경향을 분석하여 개인 맞춤형 음악추천모델을 제안한다.

II. 관련 연구

2-1. 협업적 필터링 추천시스템

협업적 필터링 추천시스템은 현재 사용자와 비슷한 취향을 가진 사용자들이 선호했던 아이템들을 사용자에게 추천해주는 것으로 메모리기반 협업필터링이라고 부른다[2]. 이 방법은 고객의 평가 점수 혹은 선호도 점수를 통해 쌓아온 아이템간의 유사성을 기반으로 추천하기 때문에 주로 온라인 쇼핑몰이나 음원사이트 등에서 사용된다. 그러나 협업필터링 방식은 초기에 신제품이나 평가가 이뤄지지 않은 제품의 경우 예측이 어렵다. 이러한 추천 방법은 사용자의 성향이 지속적으로 일관될 것이라는 전제로 이루어진다. 협업필터링에서 더 진화된 모델기반 협업필터링 추천시스템은 사전에 습득한 데이터를 가지고 예측 가능한 고객들과 유사한 성향을 가진 고객들 사이의 패턴을 파악하고 수치화하여 아이템을 추천한다[3].

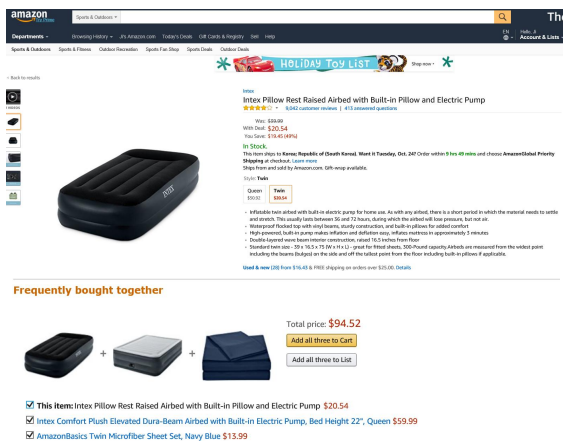


그림 1. Amazon.com 아이템 추천 예시
Fig. 1. Amazon.com Item Recommendation Example

대표적인 협업 필터링 추천서비스는 Amazon.com과 Last.fm을 들 수 있다. 그림 1과 같이 Amazon.com은 아이템 중심의 방법을 이용한다[4], [5]. 아이템 중심 방법은 아이템 간의 상관관계를 도출하여 이를 통해 사용자의 최신 데이터를 기반으로 사용자의 기호나 선호도를 예측하는 방법이다. 예를 들어 사용자가 제품을 구입한다고 가정했을 때, 다른 구매자들이 해당 제품과 함께 구입했던 아이템을 추천해주며, 비슷한 취향의 소비자들이 구매했던 비슷한 아이템을 추천해준다.

협업 필터링 추천방법이 적용된 음악 추천시스템은 Last.fm이 있다. Last.fm은 사용자들의 이용 데이터를 기반으로 음악을 추천한다. 특히 Audioscrobbler를 통해서 음악을 추천하는데, AudioScrobbler이란 Last.fm을 통해 이용자가 듣는 음악을 추적하여 자동으로 프로필에 추천음악을 추가하는 것이다[6]. 그림 2는 Last.fm의 웹페이지 화면이다.

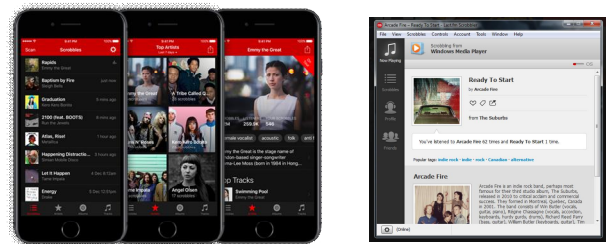


그림 2. Last.fm 모바일과 웹페이지
Fig. 2. Last.fm Mobile and Web page

Last.fm의 플러그인은 사용자의 개인 프로필을 만들고, 뮤직 플레이어 플러그인을 설치한 뒤, 노래가 절반이상 재생이 되면, 자동으로 아티스트와 노래제목이 전송이 된다. 만약 30초보다 짧게 청취가 된 곡의 경우 청취된 데이터 전송에서 제외되며, 이렇게 수집된 사용자 프로필을 통해 사용자의 음악 선호도를 예측하고, 사용자와 비슷한 취향의 사용자를 찾아서 음악 성향의 유사성을 기반으로 사용자에게 새로운 음악을 추천한다. 협업적 필터링 추천시스템의 장점은 메타 데이터베이스를 구축할 필요가 없다. 그러나 이 추천시스템의 단점은 수많은 기존 청취자의 청취데이터를 수집해야하기 때문에 시간과 비용이 많이 필요하다. 또한 많은 청취자가 선호한 인기곡 중심의 데이터 쏠림 현상 때문에 인디음악, 혹은 비대중적인 음악은 추천이 어려워진다. 그리고 새로 발표되는 곡은 기존 청취자 데이터가 존재하지 않기 때문에 실시간으로 순위가 변경되는 국내 음원 시장에서는 협업적 필터링 추천시스템을 통한 적절한 개인별 음악추천이 어렵다.

2-2. 콘텐츠 기반 추천시스템(Content-based Recommendation System)

콘텐츠 기반 추천시스템은 내용 분석을 통해 음악의 특징들을 사용자의 프로파일과 매치시켜서 평점(rating)없이도 아이템의 연관성을 가지고 선호도를 예측할 수 있는 기법이다[2].

내용 기반 추천시스템에서는 사용자의 프로파일과 상품에 대한 정보를 이용하여 추천하기 때문에 새로운 상품의 정보나 사용자 반응을 추천에 쉽게 반영할 수 있다. 대신 사용자의 수동적인 입력에 기반을 두기 때문에 사용자가 입력을 회피하거나 정보 공개를 하지 않거나, 사용자의 프로필 정보가 없는 경우에는 적절한 적용이 불가능하다[7]. 이러한 내용 기반 추천시스템을 적용한 예로는 Pandora.com가 있다.

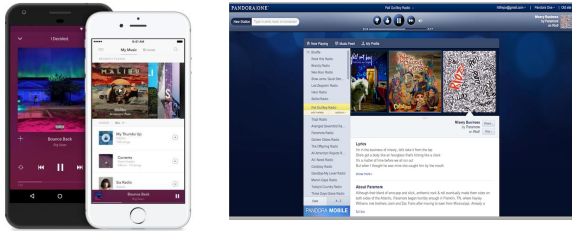


그림 3. Pandora 모바일과 웹페이지
Fig. 3. Pandora Mobile and Web page

그림 3은 판도라의 모바일과 웹페이지 화면이다. 판도라의 뮤직 게놈프로젝트는 음악 전문가들이 장르에 따라 최대 450개의 특성값을 통해 직접 분류한다. 분류된 450개의 특성을 450차원의 좌표로 변환하여 좌표간의 거리를 계산하여 노래 간의 유사도를 사용한다. 뮤직 게놈프로젝트의 450개의 분류 기준은 전체를 공개하지 않으나, 현재까지 장르, 리드보컬 성별, 그루브 사용빈도, 일렉트릭 기타의 디스토션 레벨, 백그라운드 보컬 유형 등으로 알려져 있다[8]. 판도라는 곡들의 유사도를 계산함에 있어서 단순한 거리를 사용하지 않고, 사용자의 취향을 반영하여 특성 간 다른 가중치를 두어 유클리드의 거리를 사용하여 음악을 추천한다.

개인적 음악 선호도와 음악적 특징을 반영하기 위해서 내용 기반의 음악적 분석을 통한 여러 가지 음악추천방법이 제안되었다. 원재용 외 2인은 곡의 대표 선율을 이용하여 유사한 곡을 클러스터링 하고 내용기반 검색 시 클러스터 안에 거리를 계산하여 질의와 유사한 곡을 추천해주는 기법으로 사용자의 만족도를 측정하였다[9]. Logan은 오디오데이터를 19차원의 MFCC 벡터값으로 변환하여 기준이 되는 음악세트를 만들고, 음악세트와의 거리를 측정하여 평균, 중간값, 최솟값의 음악을 추천해주는 모델을 제안하였다[10]. 김동문, 이지형은 음원의 파형 변화 분석을 통해서 사용자가 다운로드 한 파일리스트를 분석하고 벡터 유사도를 통해서 사용자의 맞춤형 추천시스템을 제안하였고[11], Li, Myaeng, Guan은 콘텐츠 분석을 통해 오디오 특징들과 사용자 선호도의 가우시안(Gaussian) 분포를 이용하여 그룹화를 하여 사용자의 음악추천 성능을 개선한 연구를 제안하였다[12].

2-3. 하이브리드 추천시스템(Hybrid Recommendation System)

하이브리드 추천시스템은 협업적 필터링 방법과 콘텐츠 기반의 추천방법을 혼합한 방법이다. Marko Balabanovic, Yoav Shoham은 Fab이라는 어플리케이션을 통해 하이브리드 접근 방법을 소개하였다. 사용자 프로필을 콘텐츠 기반 추천기법으로 관리하고, 아이템을 분석하여 유사한 사용자를 찾는 방법은 협업적 필터링 기법으로 이용하였다[13]. 추천되는 아이템들은 사용자의 프로필과 관계가 깊고, 유사한 취향을 가진 다른 사용자에게도 높게 평가될 때 추천된다. 또한 김재경 외 3인의 연구에서는 협업 필터링의 한계점을 극복하기 위해 영화와 키워드 정보를 제공하여 시뮬레이션을 실시하였고, 모바일 환경에서 베스트셀러 추천과 순수 협업 필터링에 비해 더 나은 결과를 보여 주었다[14]. 김경록, 변재희, 문남미는 모바일의 특성과 장르의 특징을 반영하여 신규 아이템이나 신규 이용자에게 추천해 줄 수 있는 방법을 제안하였다[15]. 사용자의 선호 장르 프로필과 장르 유사도 프로필을 추가하여 추천목록을 도출하였고, 이 모바일 환경에서 서비스될 수 있도록 구현한 결과, 순수 협업 필터링 추천방법보다 더 우수한 결과를 보였다. 하이브리드 추천시스템은 협업 필터링 방법과 콘텐츠 기반 추천방법을 동시에 적용하여 두 기법의 단점을 보완할 수 있으나, 다양한 프로필과 콘텐츠를 분석하기 위해서는 많은 데이터가 요구되며, 많은 시간과 비용이 소요될 수 있다.

본 연구는 콘텐츠기반 음악추천기법을 이용하여 타인의 프로파일이나 선호도에 대한 많은 기존 청취자 데이터 없이도 추천이 가능하도록 실험하였다. 1인 청취자의 음악선호여부와 음악 콘텐츠 데이터만을 통해 개인의 음악적 선호요소를 예측하고 맞춤형원을 추천해줄 수 있으며, 이러한 방법으로 많은 청취자 데이터를 수집하는데 필요한 많은 시간과 비용이 절약될 수 있다. 본 연구에서는 음악적 요소를 기반으로 사용자의 음악 선호도를 통한 개인별 맞춤형 음악추천모델을 제안한다.

III. 연구 방법

최근 5년간 연간 100위권 내에 진입한 총 460곡의 대중가요를 음악적 분석을 통해 대중음악의 전반적인 동향과 일반적인 선호 경향을 분류하였다[16]. 이를 바탕으로 개인별 곡의 선호 경향을 알아보기 위해 총 460곡의 대중가요를 음악적 특성 토대로 구분하였다. 그리고 피험자에게 가요를 들려주고 곡에 대한 선호 설문을 한 뒤, 개발된 음악추천모델에 적용하여 선형 회귀분석을 통해 곡의 선호 여부를 실험하였다.

3-1. 음악추천모델 알고리즘

총 460곡을 음악적 독립변수의 데이터 분류기준에 따라 데이터 구성하였고, Tensorflow를 이용해 선형회귀분석을 하였다. 여러 개의 독립변수를 사용함으로써 예측 정확도를 높이고자 하였다.

이 분석방법은 종속변수 y 와 독립변수인 x_1, x_2, \dots, x_k 사이의 선형관계를 나타내기 위해 사용된다. 독립변수가 k 개인 다중선형 회귀분석의 모델은 다음과 같다[17].

$$y = \text{softmax}(W_1x_1 + W_2x_2 + \dots + W_kx_k + b) \quad (1)$$

- k : 독립변수의 개수
- y : 종속변수, 사용자의 곡 선호 여부를 예측
- W_k : 선형 회귀식의 기울기, X_k 값의 가중치를 나타냄
- x_k : k 번째 독립변수의 값을 말하며, 곡의 데이터의 vector 값을 말함
- b : 편향값(bias)

본 선형 회귀분석 모델은 x_k 가 독립변수이며 y 값을 예측할 수 있도록 학습한다. x_k 값을 이용해 y 를 예측할 수 있는 최적의 W_k (가중치)와 b (편향값)을 찾도록 데이터 셋들을 1,000회 반복 훈련한다. 실험과정에서 사용자에게 곡의 선호여부의 설문결과와 예측모델 사이에 예측 가능한 확률을 알아보기 위해 예측정확도를 측정하였다.

3-2. 사용자의 곡인지 및 선호도에 따른 음악데이터 분류기준

개인별 음악 추천을 위한 모델을 설계하기 위해 음악데이터를 13개의 음악요소로 분류하였다[16]. 기존의 콘텐츠기반 음악추천기법은 BPM, 가수 등을 음향분석 데이터를 통해서 음원을 분류한다. 그러나 본 모델은 사용자가 선호하는 음악적 요소를 파악하고 음악 작곡을 할 때 결정하는 요소들을 추가하여 발행연월, 가수, 장르뿐만 아니라, 조성, 박자, Verse, Chorus의 시작시간, 못갖춘마디 여부 등을 음악선호 요소로 추가하였다. 데이터의 특징에 따라 비연속성 데이터는 원-핫 벡터(one-hot vector)로 나타냈고, 연속적인 특징을 가진 데이터는 숫자로 나타냈다. 표 1은 개인별 음악의 선호요인 분석을 위한 곡의 분석 기준을 나타낸 표이다.

표 1. 음악 분석 기준

Table 1. Music Classification standards

No	Independent Variable	Variable classification													
1	Issued year	2012			2013			2014			2015			2016	
2	Issued month	1	2	3	4	5	6	7	8	9	10	11	12		
3	Keynote	C	C#	D	D#	E	F	F#	G	G#	A	A#	B		
4	Tonality	Major						minor							
		0						1							
5	Time signature	2/2	2/4	3/4	4/4	6/8	12/8	5/4							
		1	2	3	4	5	6	7							
6	BPM	= (x-100)/100													
7	Type of Singer	Man Solo		Woman Solo		Mixed Group		Men Group		Women Group					
		1	2	3	4	5	6	7	8						
8	Music style	Dance	Ballad	R&B	Folk	Hip-hop	Indi	Pop	etc.						
		1	2	3	4	5	6	7	8						
9	OST	Released by music resources site						Released by Drama (OST)							
		0						1							
10	Incomplete bar	Complete bar						Incomplete bar							
		0						1							
11	Total play time	3minutes(180seconds) = (x-180)/180													
12	Verse	20 seconds = (x-20)/20													
13	Chorus	1minute(60 seconds) = (x-60)/60													

음악 분류기준은 총 13가지였으며, 발행연도, 발행월, 조성, 장단조여부, 박자, 곡의 빠르기, 가수의 유형, 음악 스타일, OST 여부, 못갖춘마디, 총 연주시간, Verse와 코러스의 시작시간을 계산하여 나타냈다. 비연속성 데이터는 발행년, 발행월, 조성, 장단조, 박자, 가수유형, 음악스타일, OST여부, 못갖춘마디가 있었으며, 연속성 데이터는 곡의 빠르기, 총 연주시간, Verse 시작시간, 코러스 시작시간이었다.

IV. 실험 방법

음악추천 모델을 통해 추천된 곡을 실제 피험자가 선호하는 지 실험하기 위해 20명의 피험자에게 총 2회에 걸쳐 훈련단계 실험과 검증단계 실험을 진행하였다. 훈련단계 실험은 해당 모델에 훈련 데이터 셋을 얻기 위해 시행하였으며, 검증단계 실험은 추천모델을 통해 추천된 곡의 선호여부를 확인하기 위해 실험하였다. 본 실험의 개요는 표 2과 같다.

표 2. 실험 방법 개요

Table 2. Experiment Method Overview

Participant	Eligibility : People who listen to popular music and use sound recording sites 20 participants (1: 1, twice in total, progressed to the training phase and verification phase)
Process	1) Explanation of the experiment process related to this experiment 2) Listen to the selected music Check the song's pre-recognition and preference 3) Interviews related to favorite music and music listening patterns when using a music source site

4-1. 훈련단계 실험

1차 실험은 실험 진행자와 피험자가 1대 1로 청취 질문을 통한 설문과 인터뷰를 진행하였으며, 피험자는 음악을 듣고 아는 곡인지의 여부와 선곡된 음악의 선호를 ‘좋다/싫다/모르겠다’로 답변할 수 있도록 질문지를 작성하였다. 그리고 소리 내어 말하기(Think aloud)를 실시하여 피험자가 음악 듣고 질문지를 작성하면서 필요한 질문이나 본인의 생각은 언제든지 이야기할 수 있도록 하였다. 표 3은 곡 선호여부와 선인지 여부를 측정하는 질문지 예시이다.

표 3. 훈련단계 실험 질문지

Table 3. A Training Experiment Questionnaire Example

No	Title	Singer	known		like		
1	Wildflower	Park Hyo Sin	O	X	O	X	△

질문지 답변 방식은 노트북을 통해 질문지를 작성할 수 있도록 설계하였다. 또한 피험자의 의지에 따라 곡을 끝까지 듣지 않고도 곡에 대한 답변이 끝나면 다음 곡을 들을 수 있도록 진행하였다. 그림 4는 실험용 음악 플레이어와 실험용 질문지의 예시이다.

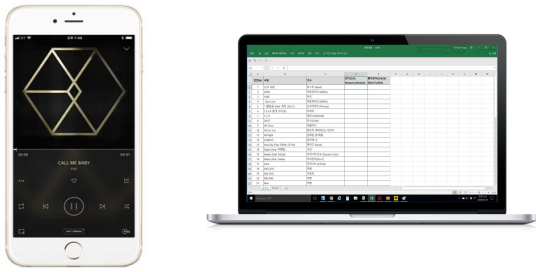


그림 4. 실험용 음악 플레이어와 질문지 예시
Fig. 4. Music player samples for experiment

피험자는 최대한 일상생활에서 음악을 듣는 방법과 같도록 스마트폰으로 음악을 재생하고, 이어폰을 통해 들리는 음악을 듣고 인지여부와 선호여부를 지필로 대답하게 하였다.

4-2 훈련단계 실험결과: 선호 곡에 대한 예측정확도

실험참가자(Participant, P) 20명의 실험을 통해 음악의 선호 여부를 조사하였으며, 해당 데이터를 음악추천모델을 통해 선호 곡에 대한 예측정확도를 도출하였다. 그림 5는 훈련단계 실험결과에 대한 그래프이다.

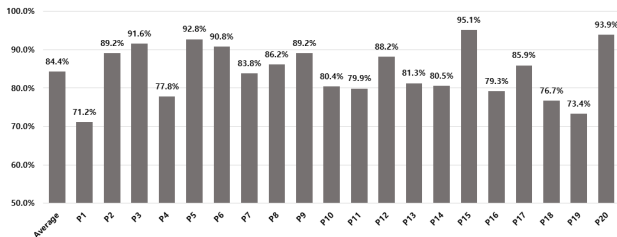


그림 5. 훈련단계 실험결과
Fig. 5. Training Experiment Results

그림 5는 훈련단계 실험에서 피험자의 선호여부를 가지고 예측정확도를 나타내었다. 전체 피험자의 평균 예측정확도는 84.4%이다. 가장 높은 예측정확도를 보여준 피험자는 P15로 95.1%를 나타냈고, 가장 낮은 예측정확도를 나타낸 피험자는 P1이었다. 전체 피험자의 예측정확도 격차는 23.9%이며, 이는 피험자의 질문지 결과를 예측했기 때문에 과적합(Overfitting)의 우려가 있다. 그러므로 개인화 추천과정을 거쳐서 선곡된 음악을 통해 피험자 검증실험을 진행하였다.

4-3. 개인화된 추천 곡의 선곡 과정

검증단계에서 개인화된 추천 곡은 다음과 같이 선별되었다. 훈련단계 실험에서 나타난 선호 곡 예측정확도를 통해 선호요소별 Parameter 값의 가중치를 계산하였다. 그리고 훈련단계 실험에서 나타난 가중치를 실험하지 않은 곡의 Parameter 값에 적용하여 Parameter 값이 높은 순으로 좋아하는 곡을 선별하며,

Parameter 값이 음숫값에서 높은 순으로 싫어하는 곡을 선별하였다. 이와 같이 개인화된 추천 곡은 좋아하는 곡만 선별할 경우 ‘좋다’로 응답이 될 수 있기 때문에 좋아하는 곡 70%와 싫어하는 곡 30%의 비율로 예측하여 선곡하였다.

4-4. 검증단계 실험

검증단계 실험에서는 피험자에게 음악추천모델을 통해 예측된 음악을 추천하였고, 추천된 곡의 선호여부를 통해 추천모델의 정확도를 확인하였다. 피험자는 검증단계 실험 후, 약 7일에서 10일 이후 동일한 장소에서 검증단계 실험에 참여하였다. 검증단계 실험 곡은 선호도 예측모델에서 높은 선호와 불호가 예측되는 곡을 선곡하여 훈련단계와 같은 방법으로 음악 어플리케이션에 피험자별 선호 곡 재생 목록을 각각 만들었으며, 피험자의 질문지는 노트북을 통해 선곡된 곡을 듣고 선호여부를 답변하도록 하였다. 검증단계 실험 질문지의 예시는 표 4와 같다.

표 4. 검증단계 실험 질문지

Table 4. A Verification Experiment Questionnaire Example

No	Title	Singer	known		like	
1	Wildflower	Park Hyo Sin	O	X	O	X

기존 훈련단계 실험 질문지는 선호도 선택을 ‘좋다 / 싫다 / 모르겠다’ 3가지로 나누었지만, 검증단계 질문지는 이미 선호도를 파악하였기 때문에 선호여부 답변에서 ‘좋다 / 싫다’의 2가지로 음악 선호도를 기입하도록 하였다.

4-5. 검증단계 실험결과: 추천음악의 예측정확도

훈련단계 실험결과를 바탕으로 음악추천모델에 의해 추천된 음악을 2차 실험결과 비교하였다. 이에 대한 실험결과는 그림 6과 같다.

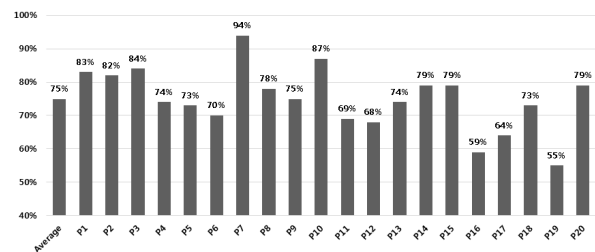


그림 6. 검증실험결과: 추천된 음악의 실험자의 예측정확도
Fig. 6. Verification Experiment Result

그림 6에서 전체 피험자의 예측정확도를 확인한 결과, 75.0%의 곡에서 예측이 일치하였다. 가장 높은 예측정확도를 나타낸 피험자는 P7이며, 94%의 높은 예측정확도를 보였다. P 가장 낮은 예측정확도의 피험자는 P19로 55%의 정확도를 보였다. 이러한 결과를 통해 훈련단계 실험을 로 피험자의 곡의 선호요인을 파악한 뒤, 음악 추천모델을 통해 음악을 추천했을

때, 20명의 피험자의 선호도 평균은 훈련단계 실험에서 예측정확도가 84.4%가 나타났으나, 검증단계 실험에서는 예측정확도가 75.0%로 약 9.6% 낮게 도출되었다. 그 이유는 훈련단계 실험은 훈련데이터를 가지고 계산된 예측정확도의 수치는 과적합의 우려가 있기 때문에 검증단계의 정확도보다 높았다. 그러나 이를 보완하고자 검증단계 실험을 통해 추천 음악이 실제로 예측이 맞았는지 예측정확도를 확인하였고, 비교적 높은 수치가 나타났다.

V. 결 론

본 연구는 국내 대중가요의 음악 분석요소를 도출하고, 사용자의 음악 선호요인을 분석하여 청취자를 위한 개인 맞춤형 음악추천모델을 제안하였다. 피험자에게 선인지와 선호여부를 질문하여 피험자 답변과 예측모델 사이에 선호 곡 예측 정확도를 도출하였다. 훈련단계 실험결과 평균 84.4%의 예측정확도가 나타났으며, 검증실험에서는 좋아하는 곡과 싫어하는 곡에 대한 예측 정확도가 평균 75%로 나타났다. 콘텐츠기반 음악추천모델로 많은 이력이나 평가점수가 없이도 추천의 가능성을 확인하였다. 본 연구는 실제 음악 분류기준을 바탕으로 개인의 음악 성향을 측정하고, 좋아하는 곡이나 싫어하는 곡의 예측이 가능하다는 것을 보여주었다. 또한 본 실험의 검증단계에서 피험자에게 실제 음악 추천을 진행하였으며, 개인화된 음악 추천 모델의 유용성을 확인하였다.

감사의 글

본 논문은 한국연구재단의 BK21 플러스 사업 (미래기반 창의인재양성형 22A20130012806) 및 여성신산업융합인재 양성사업(2015H1C3A1064579)의 지원을 받아 수행된 연구임.

참고문헌

[1] Ministry of Culture, Sports and Tourism. A Study on the Structure Analysis of Recommendation System and Analysis of Spillover Effects of Online Digital Music Distribution Company. Seoul. 2012.

[2] Adomavicius, G., Tuzhilin, A., "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, pp.734-749, April 2005.

[3] M. A. Lee, Music Recommendation System Based on Radio Episode Analysis, M.A. dissertation, Seoul National University, Seoul, 2012.

[4] Amazon.com [Internet]. Available: <https://www.amazon.com>

[5] G. Linden, B. Smith and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," *IEEE Internet Computing*, Vol. 7, No. 1, pp. 76-80, January 2003.

[6] Last.fm. [Internet]. Available: <https://www.last.fm>

[7] Pazzani, M. J., "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review*. Vol. 13. pp. 393-408. 1999.

[8] Pandora Internet Radio. *About The Music Genome Project*. Pandora Internet Radio. Vol. 17. 2004.

[9] J. Y. Won, K. I. Ku, Y. S. Kim, "Content-Based Filtering Using Representative Melody in Music Recommendation System," in *Proceeding Korean Institute of Information Scientists and Engineers*, Seoul, Vol. 31, No. 2. pp. 229-231. October 2004.

[10] B. Logan, "Music Recommendation from Song Sets," in *Proceedings of the International Conference on Music Information Retrieval*, pp. 425-528. October 2004.

[11] D. M. Kim, J. H. Lee, "User adapted music recommendation System using sound wave," *Korean Institute of Information Scientists and Engineers*, Vol. 33, No.2(B). pp. 250-253. October 2006.

[12] Q. Li, S. H. Myaeng, D. H. Guan, B. M. Kim, "A Probabilistic Model for Music Recommendation Considering Audio Features," *Information Retrieval Technology*, Vol. 3689, pp. 72-83, 2005.

[13] Marko Balabanovic, Yoav Shoham, "Content-Based, Collaborative Recommendation," *Communications of the ACM*. Vol. 40, No.3, pp.66-72. March 1997.

[14] J. K. Kim, Y. H. Cho, S. T. Kim, H. K. Kim, "A Personalized Recommender System for Mobile Commerce Applications," *Asia Pacific Journal of Information Systems*, Vol. 15, No. 3. pp. 223-241. Oct 2005.

[15] K. R. Kim, J. H. Byun, N. M. Moon, "Collaborative Filtering Design Using Genre Similarity and Preferred Genre," *Journal of the Korea society of computer and information*, Vol. 16, No.4, pp. 159-168, April 2011.

[16] J. Y. Chung, M. J. Kim, "A Study on the Trend of Korean Pop Music Preference Through Digital Music Market," *Journal of Digital Contents Society*, Vol. 18, No. 6, pp. 1025-1032, October 2017.

[17] J. Y. Chung, M. J. Kim, "Music recommendation model by analysis of listener's musical preference factor of K-pop," in *Proceedings of the 2018 International Conference on Information Science and System*, Jeju, pp. 8-11, April 2018.

정지윤(Ji-Yun Chung)



2011년 : 이화여자대학교 대학원 디지털미디어학부 (디지털미디어석사)
2018년 : 이화여자대학교 대학원 융합콘텐츠학과 (공학박사)

2011년~2013년: 한국문화예술교육진흥원 사원
2016년: 경민대학교 교양학부 강사
2015년~2018년: 이화여자대학교 융합콘텐츠학과 크리에이티브컴퓨팅랩 연구원
2018년~현 재: 제이드앤휴 크리에이티브그룹 매니저
※관심분야: 미디어 교육(e-learning), 디지털 음악, 데이터 분석, 추천시스템 등

김명준(Myoung-Jun Kim)



1991년: 한국과학기술원 (전산학 석사)
1996년: 한국과학기술원 (전산학 박사)

1996년~1997년: University of Washington Visiting scholar
1997년~2000년: 한국전자통신연구원 선임연구원
2001년~현 재: 이화여자대학교 융합콘텐츠학과 교수
※관심분야: 컴퓨터그래픽스, 영상처리, 클라우드 컴퓨팅, 데이터 분석 등