



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

음악 재생율을 활용한
음악 추천 시스템에서의
암시적 피드백 특성 발굴 연구

Exploiting music playrate
in discovering implicit feedback features
for music recommender systems

2018 년 2 월

서울대학교 융합과학기술대학원
융합과학부 디지털정보융합전공
유 수 연

초 록

디지털 음원 시장 규모가 커짐에 따라 사용자는 방대한 크기의 디지털 음원 컬렉션에 접근할 수 있게 되었으나, 동시에 그 중에서 자신이 어떤 음악을 원하는지 찾고 선택하는 행위는 더욱 어려워지고 많은 시간을 소모하게 되었다. 이와 같은 이유 때문에 음악 추천 시스템의 중요성이 부각되며 그 성능을 높이기 위한 연구들이 다양한 방법론을 통해 시도되고 있다.

추천 시스템의 목적은 ‘사용자가 소비하지 않은 아이템 중 선호/만족할만한 추천 아이템을 찾는 것’에 있으며, 특히 음악 도메인에서는 이를 위해 사용자가 어떤 음악을 어떤 기준으로 얼마나 선호/불호하였는지를 분석하여, 다음으로는 어떤 곡을 듣고 싶어 하는지, 어떤 곡을 들어야 만족도가 높을 것인지를 예측해야 한다. 이와 같은 추천 시스템의 목적으로 미루어 보았을 때 사용자의 선호도는 추천 시스템에 있어서 가장 핵심적인 요소라고 할 수 있으며, 그 동안의 추천 시스템 연구들에서는 사용자의 선호도를 모델링하기 위해 크게 명시적 피드백(explicit feedback) 과 암시적 피드백(implicit feedback) 방식을 사용해왔다. 이 중에서도 특히 암시적 피드백 방식은 사용자로부터 직접 평가를 입력받지 않아도 된다는 점에서 음악 선호도, 혹은 평가를 추정할 때 가장 큰 문제로 대두되는 희소성 문제(sparsity problem)를 보완할 수 있다는 이유로 크게 각광받고 있다.

음악 도메인에서의 암시적 피드백은 사용자의 음악 청취 기록을 통해 수집되며, 음악 추천 시스템에서는 재생/스킵/정지 등의 청취

행위로부터 얻을 수 있는 특성 중에서도 특히 특정 곡을 몇 번 들었는지를 나타내는 재생 횟수(playcount)가 대부분 사용되고 있다. 이러한 재생 횟수는 감소하지 않는(non-decreasing) 특징으로 인해 사용자의 선호도 감소를 반영하지 못하고, 고정적인 수집 기준을 가지기 때문에 사용자마다 상이할 수 있는 선호도 기준을 반영하지 못하며, 선호도 선상의 상위 극소수를 제외한 나머지 대다수 곡들에 대한 선호도는 구별하지 못한다는 한계점을 지닌다.

본 연구에서는 기존 음악 도메인에서 대표적으로 사용되는 암시적 피드백인 재생 횟수의 한계점을 보완하고 사용자의 음악 선호도를 보다 잘 반영할 수 있도록 하기 위해 기존의 연구와 사용자의 청취 기록 데이터를 기반으로 ‘가중 재생율’이라는 개념을 제안하고, 이를 바탕으로 ‘누적 가중 재생율’ 및 ‘재생횟수-평균 가중 재생율 곱’이라는 새로운 암시적 피드백 특성들을 도출한다. 또한 사용자 평가를 통해 제안된 특성이 사용자의 실제 선호도를 잘 반영할 수 있는지와 음악 추천 시스템에 적용되었을 때 성능의 차이가 있는지를 검증한다. 이 후 결과 분석을 통해 본 연구의 한계점과, 음악 선호도 모델링과 음악 추천 과정과의 관계성에 관해 고찰하고, 이를 바탕으로 연구의 결론을 도출한다.

주요어 : 음악 추천 시스템, 협업적 필터링, 음악 선호도, 암시적 피드백, 청취 기록, 재생 횟수, 재생율

학 번 : 2015-26035

목 차

제 1 장 서 론	1
제 1 절 연구 배경	1
제 2 절 연구 목적	7
제 2 장 관련 연구	8
제 1 절 이론적 배경	8
2.1.1 음악 선호도 모델	8
2.1.2 추천 시스템에서의 사용자 피드백	9
2.1.3 음악 추천 시스템의 평가	14
제 2 절 선행 연구	16
2.2.1 암시적 피드백을 활용한 추천	16
2.2.2 음악 재생을 관련 연구	19
제 3 장 연구 데이터 및 제안 특성	20
제 1 절 연구 데이터	21
3.1.1 LFM-1b 데이터셋	21
3.1.2 트랙 지속 시간 데이터 수집	22
제 2 절 제안 암시적 피드백 특성	24
3.2.1 재생을 및 가중 재생율의 정의	24
3.2.2 제안 특성 1: 누적 가중 재생율	25
3.2.3 제안 특성 2: 재생횟수-평균 가중 재생율 곱	26
제 4 장 사용자 평가	27
제 1 절 평가 과정	28
제 2 절 음악 선호도 모델 평가	30
4.2.1 평가 문항 1	30
4.2.2 평가 문항 2	32
제 3 절 음악 추천 알고리즘에서의 성능 평가	33

4.3.1	로지스틱 행렬 분해 알고리즘	34
4.3.2	데이터 샘플링	35
4.3.3	음악 추천 리스트 평가 방법	36
제 5 장 연구 결과		38
제 1 절	음악 선호도 모델 평가 결과	38
5.1.1	평가 문항 1 결과	38
5.1.2	평가 문항 2 결과	42
제 2 절	음악 추천 알고리즘에서의 성능 평가 결과	44
5.2.1	추천 곡 리스트 평가 결과	45
5.2.2	개별 추천 곡 평가 결과	47
제 3 절	결과 정리 및 고찰	54
5.3.1	선호도 모델과 추천 결과와의 관련성	55
5.3.2	음악 콘텐츠 소비 경향 차이에 따른 선호도 모델링 방식	57
제 6 장 결론		60
제 1 절	결론 및 연구 의의	60
제 2 절	연구의 한계 및 향후 연구	62
참고문헌		63
영문초록		69

그림 목차

[그림 1] 재생 횟수를 표시하는 iTunes 인터페이스	4
[그림 2] 사용자 음악 청취 행태를 재생을 벡터로 모델링한 후 군집화한 결과	4
[그림 3] 임의 사용자 평가 참여자가 청취한 곡들의 재생횟수 히스토그램	6
[그림 4] 재생횟수를 평점으로 변환하는 방식 예시	18
[그림 5] 지속 시간 데이터가 수집된 트랙들의 고유 ID값들의 분포	23
[그림 6] 가중 재생율을 구하는 방법	25
[그림 7] 전체 사용자 평가 과정	28
[그림 8] 음악 선호도 모델 평가 첫 번째 문항 구성도	31
[그림 9] 음악 선호도 모델 평가 첫 번째 문항 예시	32
[그림 10] 음악 선호도 모델 평가 두 번째 문항 예시	33
[그림 11] 음악 추천 결과 평가 문항 예시	37
[그림 12] 첫 번째 문항의 결과에 대해 통계 분석을 수행한 구조	39
[그림 13] 개별 및 전체 순위 범위에 대해 수집된 점수의 상자그림 (boxplot)	41
[그림 14] 두 번째 문항의 결과에 대해 통계 분석을 수행한 구조	43
[그림 15] 각 추천 곡 리스트별 전체 점수의 상자그림(boxplot)	46
[그림 16] 각 추천 곡 리스트별 관련성 점수의 상자그림(boxplot)	48
[그림 17] 각 추천 곡 리스트별 새로움(Novelty) 점수의 상자그림(boxplot)	51
[그림 18] 각 추천 곡 리스트별 우연성(Serendipity) 점수의 상자그림 (boxplot)	53

표 목차

[표 1] 음악정보검색 분야의 데이터 카테고리	8
[표 2] 추천 결과 평가 척도의 종류 및 설명	16
[표 3] LFM-1b 데이터셋 청취 이벤트 데이터 명세	21
[표 4] 샘플링된 LFM-1b 사용자 데이터 상세 통계	35
[표 5] 첫 번째 문항의 각 차트의 순위 범위별 점수 평균 및 표준편차와 통계 분석 결과	40
[표 6] 각 암시적 피드백 특성 기준 순위 리스트의 순위 상관분석 결과	44
[표 7] 각 추천 곡 리스트에 대한 전체 점수의 평균 및 표준편차	45
[표 8] 추천 곡 리스트 전체 점수의 대응표본 t-검정 수행 결과	47
[표 9] 각 추천 곡 리스트에 대한 관련성 점수의 평균 및 표준편차 ...	48
[표 10] 추천 곡 리스트에 대한 관련성 점수의 대응표본 t-검정 수행 결과	49
[표 11] 각 추천 곡 리스트에 대한 새로움(novelty) 점수의 평균 및 표준편차	50
[표 12] 추천 곡 리스트에 대한 새로움 점수의 대응표본 t-검정 수행 결과	52
[표 13] 각 추천 곡 리스트에 대한 우연성(serendipity) 점수의 평균 및 표준편차	52

제 1 장 서 론

제 1 절 연구 배경

국제 음반 산업 협회(IFPI)가 2017년 제공한 디지털 음원 보고서¹⁾에 따르면 세계 디지털 음악 시장의 매출 규모는 2016년 기준, 17.7%가 상승하며 약 78억 달러에 도달하였으며, 스트리밍 매출은 60.4% 상승하였고, 현재 디지털 음반 매출이 전체 음악 산업의 50%를 이루고 있다. 세계 최대의 음악 스트리밍 서비스인 Spotify²⁾에는 1억명 이상의 사용자들이 등록되어 있으며, Apple 사의 음악 스트리밍 서비스인 Apple Music 은 4천만곡 이상을 서비스하고 있다. 이러한 동향에 따라 음악 청취자들은 방대하고, 지속적으로 증가하는 크기의 디지털 음악 컬렉션에서 음악을 선택하여 들을 수 있게 되었다.

그러나 동시에 위와 같은 이유로 인해 사용자는 음악 청취 행위에 소비할 수 있는 시간에 비해 과도한 수의 곡 선택지 앞에 서게 되며, 그 중 자신이 듣고 만족할만한 음악을 찾기 위해서는 많은 노력과 시간을 소모해야 한다. 이러한 과정을 자동화한 것이 음악 추천 시스템이며, 지속적으로 증가하고 있는 음악 콘텐츠 및 소비자들과 최근 등장하고 있는 각종 인공지능 에이전트의 주요 기능 중 하나가 음악 추천 시스템임을 고려했을 때 음악 추천 시스템 연구의 필요성은 더욱 부각된다.

도메인을 막론한 일반적인 추천 시스템의 목적은 ‘사용자의 바람이나 필요를 충족시키는 아이템(item)들을 규명하는 것’에 있다(Felfernig et al., 2013). 음악 추천 시스템의 경우에는 사용자가 듣고싶어할 만한 음악 아이템을 규명하는 것을 목적으로 하며, 이를 위해서는 과거에 사용자가

1) <http://ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2017>

2) <https://www.spotify.com/>

청취한 음악 아이템 중 어떤 것을 선호 및 불호했는지 분석하는 것이 중요하다. 음악 추천 시스템은 크게 음악 내용-기반 추천(content-based recommendation), 맥락-기반 필터링 (context-based recommendation), 협업적 필터링(collaborative filtering), 하이브리드 방식(hybrid methods) 등으로 분류할 수 있다. 내용-기반 추천 방식에서는 대상 사용자가 선호했던 곡들과 음악적으로 유사한 곡을 추천하며, 맥락-기반 추천 방식에서는 사용자가 속할 수 있는 맥락과 음악과의 관계성을 분석하여 맥락과 어울리는 음악을 추천한다. 특히 협업적 필터링 추천 방식에서는 사용자와 아이템 간의 인터랙션(interaction) 행렬만을 사용하여, 벡터 유사도를 사용하여 추천 대상 사용자와 유사한 음악 취향을 가진 다른 사용자들을 찾아 그들이 청취한 음악을 추천해주거나, 행렬분해 알고리즘을 통해 행렬의 비어있는 부분에 대한 인터랙션 값을 추정하여 이를 기반으로 음악을 추천한다. 이와 같이 각 아이템에 대한 사용자의 평가(rating)를 분석한 후 이를 기반으로 추천이 이루어지는 추천 방식에서는 특정 곡에 대한 사용자의 선호도 혹은 평가를 제대로 추정하는 것이 시스템의 성능과 직결된다고 할 수 있다.

기존의 음악 추천 연구에서는 이러한 사용자의 곡 선호도 혹은 평가를 다양한 방식으로 추정해 왔으며, 이러한 방식들은 크게 명시적 피드백 (explicit feedback) 방식과 암시적 피드백 (implicit feedback) 방식으로 나눌 수 있다. 먼저 대부분의 명시적 피드백은 곡에 대한 평가를 별점 혹은 점 (하트) 등으로 표현한 형식으로 얻어진다. 사용자가 직접적으로 선호도를 표현한다는 점은 명시적 피드백의 장점 중 하나인 동시에, 직접 입력해야 하는 번거로움 때문에 전체 접근 가능한 곡 수 대비 평가된 곡의 수가 적다는 평가의 희소성 문제(sparsity problem)를 야기한다. 또한 사용자마다 평가 지표 (e.g. 별점) 에 대한 인식이 다를 수 있고 사용자가 비일관적으로 평가할 수 있다는 점에서 시스템이 사용자의 평가를 추정하는 데에 변수를 더한다(Celma, 2010).

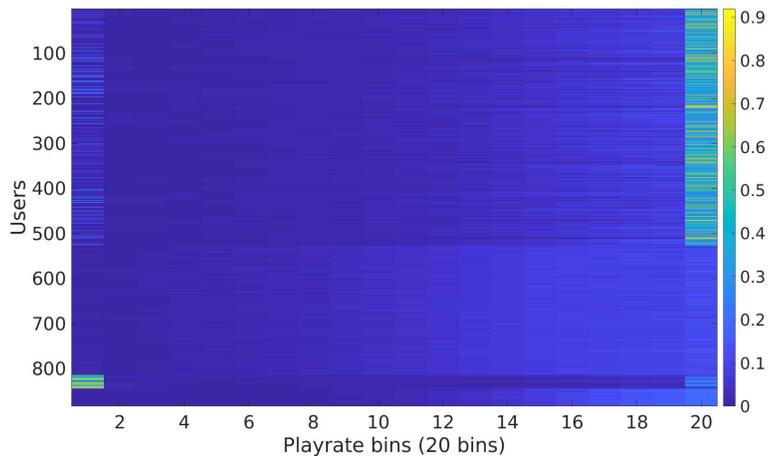
암시적 피드백 방식의 경우, 곡에 대한 사용자의 선호도를 시스템에서 자동적으로 수집하기 때문에 훨씬 적은 비용으로 많은 양의 피드백을 사용자에게 부담을 주지 않고 얻을 수 있다는 큰 장점을 가지고 있다(Song et al., 2015). 반면 단점으로는 각 아이টে에 대해 부여된 평가가 기정의 범위 내에 있지 않으며, 부정적인 피드백은 얻을 수 없다는 점이 있다. 그러나 미래의 추천 시스템에서는 사용자의 명시적인 정보 제공을 전혀 요구하지 않고, 자동으로 데이터를 수집 및 분석하여 사용자 개인의 패턴에 맞는 추천 결과를 제공하는 것이 이상적일 것이며, 이를 위해서 추천 시스템은 암시적 피드백에 더욱 의존해야 한다(Adomavicius and Tuzhilin, 2005). 이에 따라 최근의 연구 동향은 시스템-중심적 (system-centric) 인 연구에서 사용자-중심적 (user-centric) 으로 변화하고 있으며, 명시적 피드백과 암시적 피드백의 상관관계를 분석하는 연구나 사용자의 음악 선호도를 암시적 피드백을 사용해 모델링하려는 연구 등을 그 예로 들 수 있다.

음악 도메인에서 대부분의 암시적 피드백은 사용자가 어떤 곡을 청취하는지와 재생, 일시정지, 스킵, 정지 등과 같은 청취 습관으로부터 수집된다(Celma, 2010). 그 중에서도 특히 청취 기록으로부터 수집되는 곡 재생 횟수 (playcount)가 사용자의 음악 선호도를 나타내는 대표적인 암시적 피드백 특성으로 사용되고 있으며, 재생 횟수란 사용자가 어떤 곡을 반복하여 청취한 횟수를 의미한다. 일반적으로 재생 횟수는 특정 시간 이상을 들어야 재생 횟수로 인정이 되는데, 대표적인 예로 Apple 사의 멀티미디어 플레이어인 iTunes 의 경우 곡의 길이와 상관없이 전체 재생 시간의 뒤에서부터 10초 이내까지 청취해야 재생 횟수로 인정되며, 음악 추천 서비스인 Last.fm 의 경우 전체 곡 재생 시간의 절반(50%) 이상을 들어야 사용자의 취향 프로필에 곡을 재생했음을 기록하는 스크로블 (scrobble) 로 인정하도록 기본으로 설정되어 있다.

✓ Name	Time	Artist	Album	Genre	♥ Plays
✓ Derezzed	1:44	Daft Punk	Tron: Legacy	Electronica	2
✓ The Grid	1:37	Daft Punk	Tron: Legacy	Electronica	1

[그림 1] 재생 횟수를 표시하는 iTunes 인터페이스

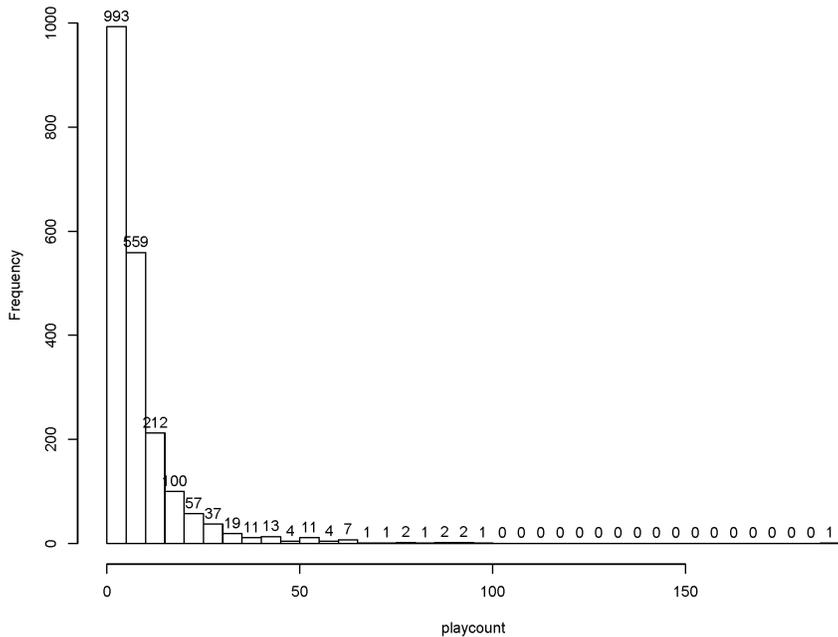
이러한 고정적인 재생 횟수 집계 방식은 모든 사용자에게 대해 같은 기준을 적용할 뿐만 아니라, 사용자마다 다르게 나타날 수 있는 청취 습관을 반영하지 못한다는 한계점을 야기한다. Yoo and Lee(Yoo and Lee, 2017)는 사용자들이 플레이백(playback) 기능을 활용하여 음악을 청취할 때 사용자들 간 청취 행태의 차이가 있는지 분석하기 위한 목적으로 사용자들의 청취 행태를 재생율 벡터로 모델링한 후 군집화(clustering)하는 연구를 시도한 바 있으며, 실험 결과를 통해 [그림 2]와 같이 사용자들의 청취 행태가 크게 네 종류의 서로 다른 특징을 가지는 군집으로 분류될 수 있음을 보였다. 이러한 결과는 재생한 길이, 혹은 스킵한 위치가 사용자들마다 다르게 해석될 수 있음을 의미하며, 사용자-의존적인(user-dependent) 방식의 암시적 피드백 수집 방식에 대한 필요성을 시사한다.



[그림2] 사용자 음악 청취 행태를 재생율 벡터로 모델링한 후 군집화한 결과

또한 재생 횟수는 위와 같은 특성 때문에 시간이 흐름에 따라 항상 증가할 수밖에 없다. 예를 들어 사용자가 어떤 곡을 들은 지 오래될수록 그 곡에 대한 재생 횟수는 높아지게 되며, 사용자가 새로 접하여 청취한 곡은 무조건 낮은 재생 횟수부터 시작하게 된다. 재생 횟수가 사용자의 선호를 암시하는 특성임을 고려해 보았을 때 앞의 상황에서는 새로 들은 곡이면 사용자가 기존에 청취하였던 곡들과 상대적으로 비교하여 무조건 낮은 선호도를 가질 것이라고 가정하는 셈이 된다. 그러나 새로운 곡은 처음부터 높은 선호도를 가질 수 없다고 가정하는 것은 직관적으로 생각해 보았을 때 타당하지 않다. 또한 Jawaheer et al. 의 연구(Jawaheer et al., 2010)에서는 사용자가 어떤 곡에 대해 명시적 피드백을 표시한 시점으로부터 전과 후의 정규화된(normalized) 재생 횟수를 비교해 보았을 때 감소함을 보인 바 있으며, 이를 참고하면 사용자의 선호도는 시간이 흐름에 따라 변화하고, 특히 감소할 수 있음을 추론할 수 있다. 그러나 앞서 설명한 것과 같이 재생 횟수는 증가하기만 하는, 혹은 감소하지 않는(non-decreasing) 특성을 가지기 때문에 이러한 사용자의 선호도 감소를 반영할 수 없다.

마지막으로 재생횟수 집계 방식은 하나의 청취 이벤트를 바이너리(binary) 값으로 결정하기 때문에, 사용자가 청취한 곡들 중 재생횟수 선상의 상위 매우 작은 비율의 곡들을 제외하고는 같은 재생횟수 값을 가지는 곡들이 대다수를 이루게 된다. 이러한 점은 사용자가 청취한 대다수의 곡들의 선호도를 분별하기 어려워진다는 문제를 야기한다. [그림 3]은 본 연구의 사용자 평가를 통해 모집된 한 참여자가 청취한 5026개의 곡들 중, 재생횟수가 1과 2인 곡들을 제외한 나머지 2038곡의 재생횟수 히스토그램을 나타낸 것이다.



[그림 3] 임의 사용자 평가 참여자가 청취한 곡들의 재생횟수 히스토그램

위 그림을 통해 극소수의 곡들을 제외하고는 낮은 재생횟수로 갈수록 같은 재생횟수 값을 가지는 곡들이 기하급수적으로 증가함을 알 수 있으며, 이러한 경향은 참여자 대부분의 재생횟수 히스토그램에서도 역시 마찬가지로 나타났다. Celma(Celma, 2010)의 연구에서는 사용자가 소비, 혹은 선호했던 아이টে에 대한 분석뿐만 아니라 소비하지 않은, 혹은 불호했던 아이টে에 대한 분석 역시 좋은 추천을 위해서는 중요하게 고려되어야 할 것임을 강조한 바 있으며, 위의 결과 역시 이와 유사한 맥락으로 선호도 선상에서 상위 극소수의 곡들뿐만 아니라 나머지 대다수의 곡들에 대한 더욱 정교한 분석이 필요함을 시사한다.

본 연구에서는 위에서 설명한 바와 같이 대표적인 암시적 피드백 특성인 재생 횟수의 고정적인 수집 기준, 증가하기만 하는 특징과 청취 이벤트의 바이너리 값으로의 맵핑 과정으로 인해 야기되는 한계점들을 보완할 수 있는 새로운 암시적 피드백 특성을 제안함으로써 사용자의 선호도를 더욱

정교하게 모델링하고, 궁극적으로는 보완된 암시적 선호도 모델을 실제 음악 추천 시스템에 적용했을 때 도출된 추천 결과에도 유의미한 차이를 이끌어내고자 하였다. 이에 따라 제 2절에서는 연구의 목적 및 연구 문제를 상세히 설명한다.

제 2 절 연구 목적

본 연구의 목적은 먼저 음악 청취 이벤트에서의 ‘재생율’ 개념을 활용하여, 음악 추천 시스템에서 암시적 피드백 특성으로 대표적으로 사용되는 재생 횟수가 가지는 한계점을 보완하면서도 사용자의 실제 음악 선호도를 유사하게 반영할 수 있는 특성들을 도출하는 데에 있다. 또한, 사용자 평가를 통해 본 연구에서 제안한 특성들이 실제 사용자의 곡에 대한 선호도를 잘 반영할 수 있는지와, 이들을 기존 협업적 음악 추천 시스템의 입력으로 사용했을 때 도출된 음악 추천 결과에 대한 사용자들의 평가에 차이가 있는지 검증하는 데에 있다. 이러한 목적에 따라 세운 세부 연구 문제는 다음과 같다.

1. 기존의 협업적 음악 추천 시스템에서 쓰이는 암시적 피드백의 한계는 무엇이며, 이를 보완할 수 있는 특성으로 무엇이 있는가?
 - 1-1. 기존의 암시적 피드백의 특징 및 한계점은 무엇인가?
 - 1-2. 제안하는 특성은 사용자 선호도와 어떤 상관관계가 있는가?
2. 제안하는 특성에 대한 평가는 어떻게 할 것인가?
 - 2-1. 제안한 특성이 기존의 한계점으로 제시하였던 점들을 보완할 수 있는가?
 - 2-2. 제안한 특성이 사용자의 실제 음악 선호도를 얼마나 잘 반영하는가?
 - 2-3. 제안한 특성을 실제 음악 추천 알고리즘에 적용하였을 때 추천 성능에 어떤 차이가 있는가?

제 2 장 관련 연구

제 1 절 이론적 배경

2.1.1. 음악 선호도 모델

사람의 음악 선호도에 대한 정확한 정의를 내리기는 어려우나, 통념적으로는 ‘음악을 좋아하는 정도’ 라고 수용되고 있다. 이러한 음악 선호도를 분석하고 세분화하려는 연구는 오래 전부터 시도되어 왔다. 음악 선호도는 크게 단기적, 장기적 선호도로 분류할 수 있는데, 각 분류는 데이터의 관점에서 서로 다르게 모델링될 수 있다. Schedl et al. (Schedl et al., 2014)은 음악정보검색(Music Information Retrieval) 분야에서는 [표 1]에 정리된 바와 같이 크게 음악 내용, 음악 맥락, 사용자 특성과 사용자 맥락 중 하나 이상의 카테고리에 해당하는 특성(feature)을 사용하거나 모델을 생성하는 방식으로 음악을 표현한다고 설명하였다.

카테고리	데이터 예시
음악 내용 (music content)	리듬, 음색, 멜로디, 화성, 크기, 가사 등
음악 맥락 (music context)	의미론적 라벨(semantic label), 연주자 명성, 앨범 커버아트, 아티스트 배경, 뮤직비디오 클립 등
사용자 특성 (user properties)	음악 선호도, 음악 훈련 정도, 음악 경험, 인구통계학적 정보, 연주자에 대한 의견 등
사용자 맥락 (user context)	기분(mood), 활동(activities), 사회적 맥락, 시·공간적 맥락, 신체적 측면 등

[표 1] 음악정보검색 분야의 데이터 카테고리

이들은 특히 사용자 맥락의 경우 대체로 역동적(dynamic)이고 자주 변화하는 요소들을 포함하면서 주로 단기적인 음악 청취 욕구에 영향을 미치며, 이와 달리 사용자 특성의 경우 사용자의 지속적(constant)이고 느리게 변하는 요소들을 포함하면서 주로 더 일반적이고 장기적인 청취 목적과 상관된다고 하였다.

대부분의 음악 추천 태스크(task)는 이상적으로 사용자의 단기적, 장기적 선호도를 모두 적절히 반영해야 좋은 추천으로 이어진다고 할 수 있다. 이 때 음악 내용, 음악 맥락 및 사용자 맥락에 해당하는 다양한 정보는 반드시 해당 정보에 대한 사용자의 음악 청취 경험에 대한 만족도와 결합되어야 의미를 가지게 된다는 점에서 사용자들이 발생시키는 방대한 양의 청취 이벤트에 대한 올바른 분석과 해석이 더욱 중요하다고 할 수 있다. 본 연구에서는 이러한 이유로 사용자의 청취 기록 중 특히 플레이백(playback) 데이터를 기반으로 가공할 수 있는 재생율(playrate) 개념을 중점적으로 활용하여 사용자의 선호도를 모델링하고자 했다.

2.1.2. 추천 시스템에서의 사용자 피드백

추천 시스템에서의 선호도 모델링은 사용자 피드백으로 이루어지며, 특히 추천 방식 중에서도 우수한 성능을 입증하고 있는 협업적 필터링의 경우 사용자 피드백이 시스템의 핵심적인 입력 요소이기 때문에 사용자 피드백을 모델링하는 것이 성능과 직결된다고 할 수 있다. 이러한 사용자 피드백은 수집 방식에 따라 크게 명시적 피드백과 암시적 피드백으로 분류된다.

명시적 피드백

추천 시스템에서의 명시적 피드백(Explicit Feedback)은 사용자가 자신이 소비한 아이템에 대해 선호도 혹은 만족도를 평점을 통해 직접적으로 표시한 것이다. 평점(rating)은 연구 도메인과는 거의 무관하게 일반적으로 N점 리커트 응답 척도(N-point Likert response scale)가 사용되어 왔으

며, 이진(binary) 평가 역시 빈번히 사용된다. 이 외에 댓글, 상품 후기 등도 명시적 피드백에 해당한다. 명시적 피드백은 사용자가 아이টে에 대해 자신의 의견을 직접 표현했기 때문에 신뢰도가 높다는 점과, 긍정적/부정적인 피드백을 모두 반영할 수 있다는 장점을 가지고 있다.

현재까지의 추천 시스템 연구는 대부분 이러한 명시적 피드백에 초점을 맞추어 왔고, 특히 응답 척도를 통해 사용자가 명시적으로 어떤 아이টে에 평가할 것을 요구하는 것이 아이টে에 대한 사용자의 관심을 표현하는 실질적인 모범 방안으로 여겨져 왔다(Jawaheer et al., 2014). 그러나 동시에 많은 연구들에서는 이러한 현재의 명시적 피드백 표시 방법이 이미 여러 한계점에 맞닥뜨려 있음을 지적한 바 있다.

먼저 명시적 피드백은 매우 사용자-침투적이다(user-intrusive). 명시적 피드백을 얻기 위해서는 사용자가 직접 아이টে에 대한 평가를 내부적으로 내린 후 시스템을 통해 이를 입력해야 하며, 이는 사용자에게 번거로운 작업이다. 이러한 사용자-침투적인 특성은 전체 아이টে에 대비 평가된 곡의 수가 적다는 희소성 문제도 야기하게 된다. 한 연구(Felfernig et al., 2013)에서는 추천 시스템의 향후 연구 쟁점으로 비간섭적 선호도 판별(Unobtrusive Preference Identification)의 중요성을 설명하며, 관련 아이টে에 사용자에게 추천하는 데에 있어서 사용자 선호도를 비침투적인 방식으로 추출하는 연구가 중요함을 강조하기도 하였다.

둘째로, 사용자들은 아이টে에 대해 명시적 피드백 혹은 평점을 제공할 때 비일관적인 태도를 보이는 경향이 있다. 자신의 실제 평가와는 다른 임의의 평점을 부여하는 사용자들도 존재하며(Lee et al., 2008), Amatriain et al. 의 연구(Amatriain et al., 2009)에서는 N점 리커트 응답 척도와 같은 방식을 사용하여 평점을 제공하게 될 경우 평점 분포 중 극단적인 점수를 제외한 중간 범위의 점수에 대해서는 사용자들이 더욱 일관성이 없으며, 사용자들의 평점을 시간이 지남에 따라 여러 번 측정하였을 때 장기간에 걸쳐 측정

할수록 비밀관성이 높아짐을 보이면서 사용자의 선호도를 측정하기에는 현재의 방식이 적합하지 않음을 보였다. 또한 추천 시스템이 사용자 평점에 의존적이기 때문에, 평점의 비밀관성이 추천의 성능에 큰 영향을 미칠 수 있다고 주장하였다. 이러한 평점의 비밀관성은 명시적 피드백의 큰 장점으로 여겨져 왔던 신뢰도에도 문제가 있을 수 있음을 암시한다.

이와 같이 추천 시스템이 현재처럼 명시적 피드백에 의존하기에는 명시적 피드백이 마주한 한계점이 많으며, 이러한 한계점들은 추천 성능에 큰 영향을 미치게 된다. 이에 따라 최근에는 명시적 피드백 대신 암시적 피드백을 추천 시스템에 적용하거나, 이보다 더 근본적인 단계에서 암시적 피드백으로 사용자의 선호도를 잘 모델링하기 위한 연구들이 증가하는 추세이다.

암시적 피드백

Kelly and Teevan(Kelly and Teevan, 2013) 의 연구에서는 암시적 피드백(Implicit Feedback)을 ‘시스템과 사용자와의 자연스러운 인터랙션을 비침투적인 방법으로 획득한 사용자와 관련된 정보’로 정의하였다. 이들은 암시적 피드백은 무엇보다 사용자가 피드백을 제공하는 비용을 필요로 하지 않으면서도 많은 양의 데이터를 얻을 수 있다는 이점을 가지고 있다는 점에서 명시적 피드백을 대체할 수 있는 매력적인 수단이라고 설명하였다. 이러한 암시적 피드백의 예로는 구매 이력, 웹 페이지에 머무른 시간, 사용자가 접속했던 웹 링크, 마우스 움직임, 미디어 플레이어 사용 (재생, 일시정지, 스킵과 정지 버튼 사용을 기록하는 것) 등이 있다(Celma, 2010). 그러나 자동으로 방대한 양의 데이터를 수집할 수 있다는 강점에도 불구하고, 암시적 피드백 역시 몇 가지 한계점을 지닌다.

첫째로, 암시적 피드백은 명시적 피드백과 달리 사용자의 부정적인 평가는 반영할 수 없다. N점 척도 등을 사용하는 명시적 평가의 경우에는 점수의 범위가 정해져 있기 때문에, 예를 들어 사용자가 중간 점수 이하로 평가하였을 경우 아이টে에 대해 부정적인 평가를 내렸다고 할 수 있으나, 암시적

피드백의 경우 시스템에서 사용자의 행동을 분석하여 긍정적인 피드백만을 기록하고 있기 때문에 사용자의 부정적인 평가는 제대로 반영할 수 없게 된다(Celma, 2010).

둘째로, 암시적 피드백은 사용자가 직접 선호를 표시한 것이 아니라는 점에서 아이টে에 대한 사용자의 의견을 제대로 반영하고 있는지가 다소 모호하다. 추천 시스템에 적용되기 위해서는 암시적 피드백이 사용자의 선호를 충분히 반영한다는 검증이 되어야 하는데, Nichols 의 연구(Nichols, 1997)에서는 사용자가 평상시처럼 읽을 때는 태도가 달라질 수 있기 때문에 암시적 데이터와 사용자의 평가간의 상관관계가 특별히 강하거나 신뢰성이 높지 않을 수 있다고 하였다. Hu et al.(Hu et al., 2008)은 사용자가 어떤 물품을 구매한 이력을 수집할 수 있다고 해도 선물을 목적으로 했거나, 구매 후에 물품이 불만족스러웠을 수도 있는 등 암시적 피드백은 본질적으로 잡음이 많은(noisy) 특성을 지닌다고 설명하였다. 또 다른 연구(Kordumova et al., 2010)에서는 추천 시스템이 일반적으로 구매 행위, 혹은 음악 서비스의 경우 청취 행위에 대해서는 긍정적인 평가로 해석하고 곡을 스킵하는 행위 등을 부정적인 평가로 해석하고 있으나, 이런 행위들을 해석하는 방법론이 명확하지 않으며 사용자마다 해석이 달라질 수도 있다고 설명하였다.

반면, 두 피드백의 추천 시스템에서의 성능이 크게 다르지 않음을 보인 연구들도 존재한다. Konstan et al. 의 연구(Konstan et al., 1997)에서는 뉴스 기사에 대해 협업적 필터링 추천 방식을 적용한 시스템을 구현함으로써 사용자들에게 뉴스 기사를 추천하였는데, 이 때 사용자들로부터 얻을 수 있는 명시적 피드백의 비율이 극히 낮기 때문에 사용자들이 기사를 읽은 시간을 암시적 평가로 사용하는 추가적인 연구를 시도하여 사용자들에게 추천할 뉴스 기사를 예측할 때 암시적 평점을 이용하는 것이 명시적 평점을 이용하는 것과 거의 유사한 성능을 보임을 검증하였다. 또한 Last.fm 의 사용자들의 명시적 피드백('Love a Track'/'Ban a Track' 기능)과 암시적 피드백(재생 횟수)을 각각 사용하여 아티스트 레벨에서 추천하는 협업적 필터링을 적용했

을 때, 두 피드백을 사용했을 때의 추천 성능이 유사함을 보인 연구도 존재한다(Jawaheer et al., 2010).

명시적 피드백과 암시적 피드백의 상관관계

실제 음악 도메인의 추천 시스템에서의 두 피드백 종류 간의 직접적인 상관관계를 분석한 연구들도 존재한다. 본 절에서 설명하는 연구들은 추천 알고리즘 자체의 성능 향상에 목적을 두기보다 사용자 선호도 모델 차이에 따른 성능 분석, 혹은 선호도 모델 자체의 비교에 초점을 맞추었다.

Jawaheer et al.의 연구(Jawaheer et al., 2010)에서는 Last.fm 사용자들의 명시적 피드백과 암시적 피드백을 비교하는 연구를 진행하였다. 이들의 연구에 의하면 먼저 시간이 지날수록 사용자가 명시적 피드백을 제공하는 비율은 감소하였다. 또한 이들은 사용자가 특정 곡에 대해 명시적 피드백 (Last.fm 의 ‘Love a Track’ 기능을 의미함)을 표시하기 전과 후의 영향을 비교하였으며, 실험 결과 곡에 대해 명시적 피드백을 제공하기 전보다 후에 재생 횟수가 감소한 사용자가 전체 사용자의 3분의 2를 차지했다. 이를 통해 명시적 피드백을 제공하는 것이 사용자의 청취 경험에 부정적인 영향을 미침을 알 수 있으며, 추가적으로 현존하는 데이터셋을 통해 얻을 수 있는 사용자들의 명시적 피드백 데이터는 연구 목적으로 그대로 사용하기에는 정확성 및 신뢰도가 낮을 수 있다는 점을 시사한다.

또 다른 연구(Parra and Amatriain, 2011)에서는 음악 도메인에서의 명시적 피드백과 암시적 피드백의 관계를 분석함으로써 암시적 피드백만이 주어졌을 때 알려지지 않은 사용자의 평점을 추정할 수 있는 선형 모델을 형성하기 위해 Last.fm 사용자들을 대상으로 설문을 진행하였다. 실험은 사용자들의 재생 목록 내에 있는 음악 앨범에 대한 평가를 5점 척도의 별점으로 매기도록 요청하는 방식으로 진행되었으며, 암시적 피드백 중 재생 횟수, 인기도(popularity)와 최근성(recentness)을 변인으로 하여 이러한 요인들과 수집된 명시적 피드백과의 상관관계를 분석하였다. 연구 결과에 의하면 명시적

피드백은 재생 횟수 및 최근성과 큰 상관관계가 있었다.

2.1.3. 음악 추천 시스템의 평가

Aggarwal의 연구(Aggarwal, 2016)에서는 음악 추천 시스템의 평가는 크게 사용자 스터디, 온라인 평가, 오프라인 평가의 세 가지 방법으로 나뉘어 이루어질 수 있다고 하였다. 먼저 사용자 스터디의 경우, 피험자를 모집한 후 이들에게 추천 시스템을 사용하여 특정 태스크를 수행하도록 지시하거나, 시스템을 통해 추천된 아이템에 대한 평점(rating)을 수집하는 등의 방식으로 시스템에 대한 평가가 이루어진다. 이러한 방식은 사용자와 시스템 간의 직접적인 인터랙션 정보를 수집할 수 있게 한다는 점에서 강점을 가지나, 피험자 모집 과정의 어려움과 경제적 비용에 비해 평가 결과에 영향을 미칠 수 있는 변수들이 많아 결과를 높이 신뢰하기가 어렵다는 단점을 가진다.

온라인 평가는 사용자 평가와 유사하나, 완성되었거나 이미 상용화되어 있는 시스템을 사용 중인 사용자들의 알고리즘 전환율(conversion rate)이나, A/B 테스트³⁾ 등의 방식을 통해 시스템을 평가하는 방식이다. 이 방식은 기존에 시스템을 이미 사용 중인 사용자를 피험자로 모집한다는 점에서 실험 결과에 편향(bias)을 생성할 수 있는 변수들로부터 비교적 안정적이거나, 이미 완성된 시스템이 아닌 초기 단계의 시스템의 경우에는 이러한 온라인 평가 방식을 적용하는 것이 어렵다.

오프라인 평가의 경우, 공개 데이터셋의 평점 데이터를 분할하여 일부는 학습 데이터, 일부는 실측값(ground truth)로 사용하여 추천 알고리즘을 통해 추정된 평점과 실측값과의 오차(error)를 최소화하는 방식으로 주로 평가가

3) 사용자 경험 디자인과 같은 온라인 영역에서, 가설을 검증하기 위해 변수 A와 변수 B에 대해 사용자가 보이는 응답을 분석하여 두 변수 중 어느 것이 더 효과적인지를 판별하는 기법

이루어진다. 오프라인 평가 방법은 사용자의 개입 없이 데이터만으로 평가하는 방식이기 때문에 반복적인 알고리즘 튜닝을 통해 알고리즘의 성능을 평가하고 개선할 수 있다는 점에서 편의성이 높으며, 오프라인 평가를 위한 평가 프레임워크 및 평가 지표들이 다수 개발되어 있기 때문에 추천 시스템의 평가는 대부분 오프라인 방식으로 이루어지고 있다. 그러나 동시에, 오프라인 평가에 사용되는 데이터셋은 그 수가 한정적일뿐만 아니라 주기적으로 갱신되는 것이 아니기 때문에 추천 알고리즘을 평가할 당시의 결과가 미래에도 좋을 결과일 것임을 보장할 수 없으며, 추천 결과에 대한 실제 사용자 입장에서의 다양한 기준(참신성, 우연성 등)에 대해서는 평가할 수 없다는 단점을 가진다.

위에서 서술한 바와 같이 대부분의 연구에서 추천 시스템의 평가를 위해 오프라인 방식을 택하고 있으며, 이 때문에 추천 시스템의 성능은 사용자의 평점 데이터와의 오차를 줄이는 목적으로 하는 정확성(accuracy) 척도로 설명되는 경우가 많다. 그러나 최근 몇 년 간 증가한 연구들은 사용자가 인지하는 ‘좋은 추천’이란 정확성만으로 평가될 수 없기 때문에, 정확성 이외에도 다양한 척도를 통해 추천 시스템의 성능을 평가하고, 상황 및 사용자들의 특성에 따라 서로 다른 척도가 추천 시스템의 목적이 될 수 있음을 설명해왔다. 최근 한 연구에서는 이러한 척도들의 종류 및 관련 연구를 조망한 후, 특정 척도를 최적화의 목표로 삼는 것이 다른 척도에 어떠한 영향을 미치는지 연구한 바 있다(Kaminskas and Bridge, 2016). [표 2]는 기존의 연구들에서 제안된 추천 결과 평가 척도의 종류 및 각 척도가 추천 결과에 대해 측정하고자 하는 측면을 요약하여 설명한 것이다.

척도 구분	설명
정확성 (Accuracy)	아이템에 대해 추정된 평점이 해당 아이템에 대한 실제 사용자의 평점과 얼마나 유사한가?
관련성 (Relevance)	추천된 아이템이 사용자가 과거 소비한 아이템과 비교하여 얼마나 관련이 있는가?
다양성 (Diversity)	얼마나 다양한 분야(장르, 주제, 특성 등)의 아이템들이 추천되었는가?
우연성 (Serendipity)	예상하지 못했으면서도(unexpected) 동시에 기분 좋은(pleasant) 아이템이 추천되었는가?
새로움 (Novelty)	사용자가 이전에 보거나 소비한 적 없는(unseen) 새로운 아이템이 추천되었는가?
커버리지 (Coverage)	전체 사용자들에게 추천된 독립적인 아이템들의 수가 전체 아이템 카탈로그에서 얼마만큼의 비율을 차지하는가?

[표 2] 추천 결과 평가 척도의 종류 및 설명

음악 도메인의 추천 시스템에서도 역시 이러한 기준은 동일하게 적용될 수 있다. 본 연구에서는 제안한 암시적 피드백 특성의 선호도 모델로써의 성능과 추천 알고리즘에서의 성능을 평가하기 위해 사용자 스터디 방식을 택하였으며, 평가 결과의 다양한 해석을 위해 사용자의 만족도 점수 이외에도 관련성, 우연성, 새로움 척도 역시 평가 기준으로 택하였다.

제 2 절 선행 연구

2.2.1. 암시적 피드백을 활용한 추천

앞에서 설명한 사용자-아이템 행렬의 항목 값으로 사용자의 명시적 평점 대신 암시적 피드백 역시 사용될 수 있다. 그 이유는 2.1절에서 설명

한 바와 같이, 비록 모델링 방식의 해석과 적용이 연구마다 제각각일지라도 암시적 피드백도 많은 경우 사용자의 선호도라고 해석되고 있기 때문이다. 사용자의 음악 청취 기록으로부터 얻을 수 있는 데이터를 암시적 피드백으로 사용하여 협업적 음악 추천 시스템을 구축한 연구는 다수 존재한다.

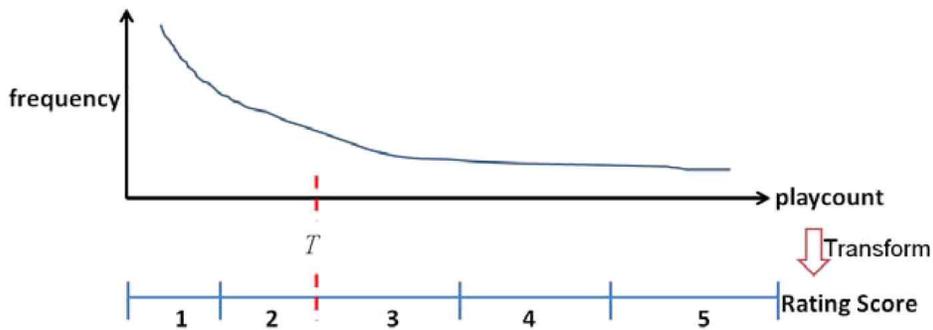
Cao and Guo(Cao and Guo, 2008) 는 기존의 곡 레벨에서의 추천 방식에 아티스트 레벨의 조건까지 더하는 추천 시스템을 제안하였는데, 이 때 사용자의 선호도 모델로 암시적 피드백을 사용하였다. 특히 이들은 암시적 피드백을 추출할 때 특정 곡을 듣는 각 청취 이벤트가 모두 같은 가중치를 갖는 것이 합리적이지 않으며, 사용자가 같은 곡을 듣는 횟수가 증가할수록 암시적 피드백에 영향을 더하는 것은 맞지만 점점 적은 가중치로 더해지는 감퇴 현상(decay phenomenon)이 일어난다고 주장하였다. 이들이 제안한 각 청취 행위에 대한 가중치는 다음과 같이 구해진다. 식에서 α 는 감퇴 비율을, β 는 암시적 피드백의 범위를 조절하는 변수이다.

$$weight(rank) = \beta \times e^{-\alpha \times rank}$$

이 후, 어떤 곡에 대한 평가는 그 곡을 들었던 청취 행위들의 가중치의 합으로 추정할 수 있으며, 이렇게 구해진 곡들의 평가를 사용자-아이템 항목 값으로 사용하여 기존의 협업적 필터링 방식으로 상위 N곡을 추천하는 시스템을 구현하였다. 사용자들의 청취 행위를 기존의 방식과 다르게 해석하여 곡들의 암시적 피드백을 모델링했다는 점은 이 연구의 차별점 중 하나이나, 이 후 등장한 몇몇 연구에서 밝혀진 바에 의하면 사용자들의 청취 이벤트는 최근에 발생했을수록 신뢰도, 혹은 가중치가 높았다는 점과, 제안한 피드백 모델을 사용하여 도출된 추천 결과에 대해 제대로 평가를 하지 않았다는 점에서 한계가 존재한다.

사용자 평가의 가장 큰 문제점으로 여겨지는 평가의 비밀관성 문제를 해결하기 위해 소셜 미디어 태그를 이용한 연구(Su et al., 2013)에서는

아이템 태그, 아티스트 태그와 재생 횟수를 사용하여 사용자의 선호도를 모델링하였다. 특히 이들은 재생 횟수가 사용자의 선호도를 나타낼 수 있다고 설명하며, [그림 4]와 같은 분포에 따라 재생 횟수를 사용자의 평점으로 맵핑(mapping) 하였다. 다음으로 이를 사용자-아이템 행렬값으로 사용하고, 일반적으로 이웃-기반 추천 시스템에서 이웃 간의 유사도 계산을 위해 사용되는 피어슨 상관계수(Pearson Correlation Coefficient) 대신 태그 빈도수를 기반으로 한 유사도를 사용한 협업적 추천 시스템을 구현하였다.



[그림 4] 재생횟수를 평점으로 변환하는 방식 예시(Su et al., 2013)

이 연구에서도 역시 자체적인 변환 방식을 구상하여 암시적 피드백인 재생 횟수를 일반적인 명시적 피드백 척도로 변환하였으나, 본 연구에서 지적하고 밝힌 바와 같이 특정 곡의 재생횟수가 쌓였다고 해도 사용자의 선호도는 감소할 수 있으며, 재생횟수는 시간-의존적인 특성인 반면 5점 척도의 점수는 시간-독립적인 특성이라는 점에서 이와 같은 변환 방식의 적절성에 의문이 남는다.

앞서 서술된 연구들을 비롯하여, 암시적 피드백을 활용한 음악 추천 연구들은 가장 근본적인 레벨의 사용자 선호도를 모델링하는 데에 대부분 재생횟수를 활용하고 있으며, 재생횟수를 사용하여 사용자 선호도, 혹은 평점 모델을 형성하는 과정 역시 연구자에 따라 다양하게 해석 및 정의되고 있다. 이와 같이 암시적 피드백 모델에 대한 해석이 수렴하지 않고 있는 실정에서,

음악 재생 시스템이 제공하는 여러 플레이백 기능을 기반으로 더욱 다양한 데이터 특성들을 가공하고 사용자의 청취 행태 및 선호도와와의 관계를 분석할 수 있는 여지가 많음에도 불구하고 플레이백 기능을 기반으로 곡에 대한 선호도를 모델링하고자 하는 연구는 음악 추천 연구 분야 중 매우 작은 비율을 차지한다.

2.2.2. 음악 재생을 관련 연구

위와 같은 실정에도 불구하고, 특히 ‘재생율’의 개념을 활용하여 음악 청취 이벤트를 모델링한 연구도 소수 존재한다. Kordumova et al.(Kordumova et al., 2010)은 ‘재생율’의 개념을 포함한 여러 청취 이벤트 기반의 특성을 사용하여 음악의 선호 여부를 추정하는 연구를 진행한 바 있다. 이 연구에서는 재생 횟수, 곡이 평균적으로 재생된 비율(재생율), 재생 시간대, 재생 날짜, 재생 요일, 스킵 횟수 등의 청취 행태상의 암시적 특성들을 사용하여 사용자들의 곡에 대한 선호(like) 및 불호(dislike)를 예측하기 위해 베이지안 분류(Bayesian classifier) 기법을 사용하였다. 이 후 사용자들의 명시적 평가 데이터를 정답(ground truth)으로 사용하여 여러 특성 조합별 정확도를 평가했을 때 평균적으로 약 76%의 정확도를 달성했으며, 이에 따라 청취 행태상의 암시적 특성을 사용하여 명시적 평가를 추정할 수 있는 가능성을 시사하였다. 또한, 다양한 특성 조합에 대한 정확도 결과가 사용자 구성에 따라 달라질 수 있음을 보이면서 사용자-독립적인(user-independent) 고정된 행태적 특성을 사용하는 것보다는 더욱 복잡한 암시적 추론 구조가 필요함을 강조하기도 하였다.

최근의 연구에서는 본 연구에서 주장하는 바와 같이 사용자들의 소비 행위로부터 부정적인 평가(rating)을 추론하는 방법론에 대한 연구가 부족함을 지적하였으며, 이에 따라 사용자들이 음악을 실제로 청취한 시간과 반복재생 여부를 고려하여 암시적 데이터를 필터링하는 방법을 제안하였다(Hanna, 2017). 이를 위해 먼저 트랙의 지속 시간 데이터를 포함하여 음악 스트리밍

서비스 사용자들의 청취 이벤트가 수록된 자체적인 데이터셋을 구축하였으며, 자체적인 기준을 설계하여 청취 이벤트를 스킵, 스트림(stream), 반복재생(replay)로 분류, 청취한 곡을 암시적 선호(implicit like) 및 암시적 불호(implicit dislike)로 분류하였다. 이 후, 필터링되지 않은 암시적 피드백 데이터, 30초 미만으로 청취한 곡들을 필터링한 데이터와 암시적 선호로 추정되는 곡들만을 포함하는 데이터의 총 세 가지 입력을 사용하여 기존 추천 알고리즘에서의 성능을 검증했으며, 평가를 통해 추천 알고리즘의 입력으로 사용하는 데이터를 청취 시간과 반복 재생 데이터를 기준으로 필터링하는 것이 사용자가 암시적으로 선호했던 곡을 추천하는 비율을 작은 수준 높이는 반면, 암시적으로 불호했던 곡을 추천하는 비율은 상당 수준 낮출 수 있음을 보인 바 있다.

전자의 연구는 제안한 특성들을 사용하여 선호 및 불호의 바이너리(binary) 값으로의 분류를 시도하였고, 후자의 연구에서는 재생을 기반의 다양한 기준으로 청취 이벤트를 사전에 필터링하는 것이 추천의 성능에 어느 정도 영향을 미침을 검증했다. 본 연구에서는 기존의 암시적 피드백의 특징과 한계를 분석하고, 이를 보완하기 위해 재생을 개념을 새롭게 활용하여 여러 암시적 피드백 특성을 제안한 후, 이들이 실제 사용자들의 선호도를 잘 표현할 수 있는지와 추천 알고리즘에 적용하여 도출된 추천 결과에 대한 성능을 평가했다는 점에서 기존 연구와 차이점이 있다고 할 수 있다.

제 3 장 연구 데이터 및 제안 특성

본 장에서는 먼저 연구 수행을 위해 선정된 데이터셋 및 추가적인 데이터 수집 과정에 대해 설명하고, 본 연구를 통해 제안된 암시적 피드백 특성에 관해 자세히 서술한다.

제 1 절 연구 데이터

3.1.1. LFM-1b 데이터셋

1장에서 설명한 바와 같이 본 연구의 목적은 음악 청취 시의 재생율이라는 개념을 활용하여 암시적 피드백 특성을 발굴하는 데에 있으며, 이를 위해서는 곡의 전체 지속 시간 데이터와, 청취자가 특정 곡을 어디까지 청취했는지를 나타내는 데이터를 포함한 충분한 양의 음악 청취 기록 데이터가 필요하다. 이와 유사한 데이터를 제공하는 데이터셋으로는 Million Song Dataset(Bertin-Mahieux et al., 2011), Celma 의 데이터셋(Celma, 2010)과 Yahoo! 데이터셋(Dror et al., 2012) 등이 있으나, 이들은 청취 기록을 타임스탬프(timestamp) 단위로 기록하지 않았거나, 데이터가 암호화되어 추가적인 데이터 수집이 불가능한 점 등의 이유로 선정되지 않았다.

LFM-1b 데이터셋(Schedl, 2016)은 사용자-중심적이고 멀티모달(multimodal)한 음악 검색 및 추천 연구의 발전을 위해 Schedl이 2016년 공개한 데이터셋으로, 음악 추천 서비스인 Last.fm⁴⁾ 을 통해 2013년 1월부터 2014년 8월까지 사용자 약 12만명이 청취한 청취 이벤트 10억 개 이상을 포함하고 있다. 청취 이벤트 데이터의 명세는 다음과 같다.

사용자 ID	아티스트 ID	앨범 ID	트랙 ID	곡을 청취한 시간 (timestamp)
--------	---------	-------	-------	--------------------------

[표 3] LFM-1b 데이터셋 청취 이벤트 데이터 명세

청취 이벤트에 등장한 아티스트, 앨범 및 트랙의 문자열 이름은 모두 별도

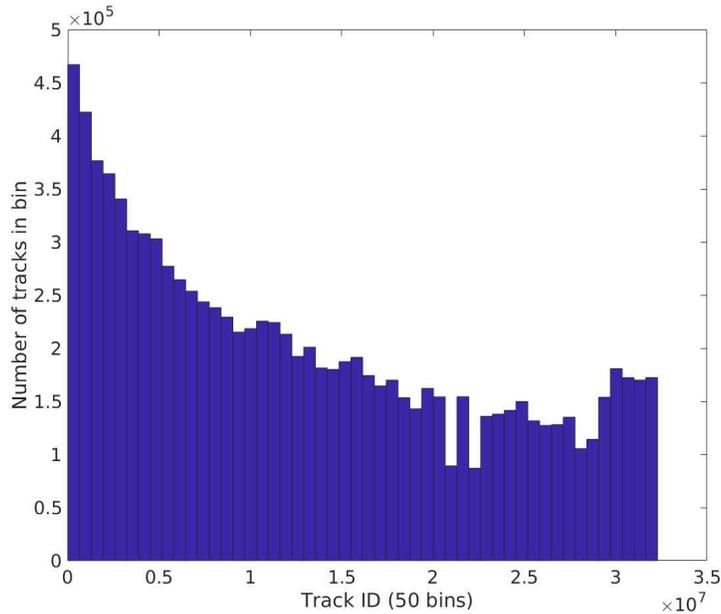
4) 영국에 기반을 둔 음악 서비스로, 사용자들의 청취 기록을 분석하여 음악 취향 프로필을 형성한 후 이를 바탕으로 사용자의 취향에 맞는 곡을 추천하고, 비슷한 취향의 사용자들을 연결시켜주는 등의 기능을 제공한다

의 파일로 기록되어 있으며, 이 외에도 사용자에 대한 간략한 인구통계학적 특징(거주국가, 성별, 나이 등), 사용자가 새로운(novel), 혹은 대중적인(mainstream) 곡을 청취하는 경향의 정도, 음악을 청취한 이벤트 수의 요일별 평균 등 사용자의 음악 청취 특성과 관련된 데이터 역시 포함하고 있다.

위 데이터셋에 기록된 청취 이벤트에 대해 인접한 청취 이벤트와의 타임스탬프 차이를 구하면 사용자가 해당 트랙을 얼마나 들었는지 알 수 있으나, 데이터셋에는 트랙 지속 시간 데이터가 기록되어 있지 않다. 이에 따라 데이터셋에 포함된 모든 트랙에 대해 지속 시간 데이터를 수집하는 과정을 추가적으로 수행하였다.

3.1.2. 트랙 지속 시간 데이터 수집

트랙의 지속 시간(duration)이라 함은 곡의 고유 길이를 의미하며, 본 논문에서 표기하는 지속 시간은 모두 초(second) 단위이다. 연구 데이터 마련을 위해 LFM-1b 데이터셋에 포함된 모든 트랙에 대해, Last.fm 및 Spotify에서 제공하는 API(Application Programming Interface) 함수 중 트랙의 정보를 제공하는 함수의 입력으로 트랙의 제목 및 해당 트랙의 아티스트명을 사용하여 트랙 지속 시간 데이터를 수집하였다. 데이터셋 내 전체 32,291,134개의 트랙의 약 32%인 10,237,395개의 트랙에 대해 지속 시간 데이터가 수집되었으며, 이들을 제외한 트랙들의 경우 수집 당시 데이터베이스에 지속 시간이 0으로 기록되어 있거나, 해당 트랙이 데이터베이스에 존재하지 않거나, 트랙 및 아티스트명의 파싱(parsing) 문제로 검색되지 않는 등의 이유로 지속 시간 데이터가 수집되지 않았다. 지속 시간 데이터가 수집되지 않은 곡들에 대해서는 본 연구에서 사용 및 제안하고자 하는 암시적 피드백 특성을 가공할 수 없다는 문제가 있으나, 다음과 같은 이유로 전체 트랙 수 대비 지속 시간 데이터가 수집된 32%의 트랙을 사용하여 연구를 진행하기에 무리가 없다고 판단하였다.



[그림 5] 지속 시간 데이터가 수집된 트랙들의 고유 ID값들의 분포

먼저 LFM-1b 데이터셋의 트랙은 1-32291134 까지의 정수형 고유 ID값을 가지며, 데이터셋 저자로부터 트랙들이 어느 정도의 대중성(popularity)을 내림차순으로 반영한 순서로 기록되어 있음을 확인하였다. 이는 고유 ID값이 작을수록 대중성이 높은 경향이 있음을 의미하는데, [그림 5]에서 볼 수 있듯이 지속 시간이 수집된 약 천만 개의 트랙의 고유 ID값들의 분포를 그렸을 때 실제로 ID값이 작은 트랙들에 대해 지속 시간 데이터가 더 높은 비율로 수집되었음을 확인할 수 있다. 이러한 수집 경향과, 앨범 매출 및 디지털 음원 재생 등을 포함한 음악 소비가 극소수의 대중적인 아티스트 혹은 곡에 편향되어 있는 현상(Celma, 2008)을 고려하였을 때, 추후 사용자평가를 위해 샘플링한 데이터셋 내의 사용자 약 1600명이 청취한 곡은 대부분 앞서 언급된 천만 트랙 이내에 포함될 가능성이 높다고 할 수 있다.

제 2 절 제안 암시적 피드백 특성

3.2.1. 재생율 및 가중 재생율의 정의

본 연구에서는 암시적 피드백 특성 발굴을 위해 음악 청취 시의 ‘재생율(playrate)’의 개념을 사용하였으며, 재생율은 ‘곡(트랙)의 지속 시간(duration) 대비 사용자가 해당 곡을 청취한 이벤트의 재생 시간’으로 정의하였다. 이러한 재생율 개념의 도입은 재생횟수가 하나의 청취 이벤트를 모든 사용자에게 대해 같은 기준을 두고 바이너리(binary) 값으로 결정하는 것과 달리 작은 단위로 사용자의 청취 행위를 기록함으로써 부정적인 청취 행위를 기록하고, 사용자가 청취한 곡들에 대해 사용자-의존적인(user-dependent) 기준이 되는 것을 가능하게 한다. 또한, 더 작은 단위를 사용함에 따라 보다 많은 곡들을 선호도 선상에서 구별할 수 있게 되며, 이를 통해 최상위의 좁은 구간에 해당하는 곡들뿐만 아니라, 중위 및 하위 순위의 곡들에 대해서도 정교한 선호도를 추정할 수 있게 된다.

한편, 암시적 피드백 데이터 역시 잡음이 많은(noisy) 특성을 가지고 있다. 사용자들은 처한 맥락에 따라 음악 플레이어와의 활발한 인터랙션이 어려운 경우 곡에 대한 자신의 실제 선호도 혹은 청취 욕구와 관계없이 계속해서 곡을 끝까지 듣는 경우도 종종 있으며, 순간에 청취하고 싶은 곡을 찾아가는(단기 선호도를 충족시키기 위한) 과정에서 실제로는 장기적으로 보았을 때 선호하는 곡을 스킵하기도 한다. 본 연구에서는 이러한 잡음(noise) 데이터의 영향을 줄이고자 일반 재생율 대신 ‘가중 재생율(weighted playrate)’을 제안하였다. 가중 재생율은 사용자가 특정 곡을 과거에 청취해왔던 기록의 신뢰도가 높다고 가정하여, 그 곡에 대한 청취 이벤트가 발생할 때 기존에 구해진 암시적 피드백에 높은 가중치를, 계산 대상인 청취 이벤트에는 낮은 가중치를 주어 이들의 합을 구한 값이며, [그림 6]은 가중 재생율을 구하는 방식을 도식화한 것이다.



[그림 6] 가중 재생율을 구하는 방법

위의 설명에 따라 어떠한 곡 t 에 대해 i 번째로 발생한 청취 이벤트의 가중 재생율을 구하는 식은 (1)과 같다. 이 때, α 는 과거 청취 기록에 부여하는 신뢰도에 따른 0-1 사이의 가중치 값이며, 본 연구에서는 데이터 관찰 후 이 값을 0.9로 선정하였다.

$$WPR(Weighted\ Playrate)_{t,i} = \alpha \times Playrate_{t,i} + (1 - \alpha) \times WPR_{t,i-1} \quad (1)$$

3.2.2. 제안 특성 1: 누적 가중 재생율

앞서 설명한 가중 재생율의 개념을 활용하여 첫 번째로 제안하는 암시적 피드백 특성은 ‘누적 가중 재생율(Cumulative Weighted Playrate)’이다. 누적 가중 재생율은 ‘특정 곡에 대해 발생한 사용자의 청취 이벤트들의 가중 재생율 값들의 합’으로 정의할 수 있으며, 이를 구하는 식은 다음과 같다. 아래의 식에서 k 는 t 에 대해 발생한 사용자의 청취 이벤트의 개수를 의미한다.

$$CWPR_t = \sum_{i=1}^k WPR_{t,i} \quad (2)$$

누적 가중 재생율은 기존의 재생횟수와 유사하게 지속적으로 누적되는 값이나, 바이너리 값이 아닌 가중 재생율만큼 누적된다는 점에서 재생횟수에 비해 보다 많은 곡들을 구별할 수 있게 하며, 선호도 선상에서 과소평가(재생

횟수 0으로 기록) 혹은 과대평가(재생횟수 1로 기록)되는 청취 이벤트의 수를 줄일 수 있을 것으로 예상할 수 있다.

3.2.3. 제안 특성 2: 재생횟수-평균 가중 재생율 곱

두 번째로 제안하는 특성은 ‘재생횟수-평균 가중 재생율 곱 (log(Playcount)*Average Weighted Playrate)’ 이다. 이 특성은 사용자가 특정 곡을 평균적으로 어느 정도까지 청취했는지가 선호도 선상에서 중요한 역할을 할 것이라는 가정 하에 고안되었으며, 이에 따라 재생횟수-평균 가중 재생율 곱 특성은 재생횟수에 자연로그를 적용한 값을 신뢰도 가중치로 사용하여 평균 가중 재생율에 곱함으로써 구해진다. 이를 구하는 식은 다음과 같다.

$$(\text{playcount} \times A WPR)_t = \ln(\text{playcount}_t + 1) \times \frac{\sum_{i=1}^k WPR_{t,i}}{k} \quad (3)$$

이 특성은 기존의 재생횟수와 앞서 제안한 누적 가중 재생율 특성과는 달리 감소할 수 있다는 특징을 가진다. 이에 따라 사용자가 특정 곡을 여러 번 청취했다고 하더라도, 시간이 지남에 따라 해당 곡을 스킵하는 행위의 빈도가 증가하면 해당 곡에 대한 암시적 피드백 값은 감소하게 된다. 제안한 두 번째 특성을 통해, 곡에 대해 누적된 선호도를 어느 정도 반영하면서도 해당 곡에 대한 부정적인 피드백을 적극적으로 반영함으로써 궁극적으로는 곡의 후반부까지 청취한 이벤트가 많으면서, 곡의 초반부에서 스킵한 이벤트가 적을수록 높은 값을 갖도록 유도하였다.

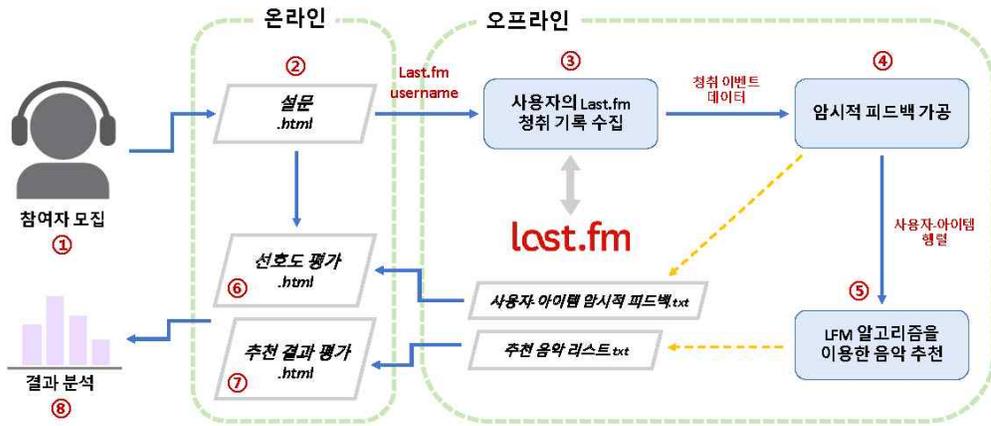
제 4 장 사용자 평가

본 연구에서는 사용자들의 곡에 대한 선호도를 잘 추정할 수 있는 암시적 피드백 특성의 발굴 및 이에 대한 평가를 목적으로 하였고, 이를 고려했을 때 기존의 암시적 피드백 특성은 평가 기준, 혹은 정답으로 사용할 수 없다. LFM-1b 데이터셋의 사용자들이 청취한 곡들에 대해 명시적 피드백 데이터를 수집하여 제안한 암시적 피드백 특성을 이와 비교하는 방법도 고려할 수 있으나, 기존 연구를 통해서도 지적된 바가 있듯 명시적 피드백 데이터는 근본적으로 잡음이 매우 많고 비일관적이기 때문에 이 역시 신뢰할만한 평가 기준이 되지 못한다.

이에 따라 본 연구에서는 실제 사용자의 개입을 통해 제안된 암시적 피드백 특성에 대한 평가를 수행하는 것이 필수적이라고 판단하여 사용자 평가를 수행하였다. 그러나 암시적 피드백 특성이 선호도를 잘 반영하는지 직접적으로 평가할 수 있는 기존의 프레임워크가 존재하지 않아, 자체적인 평가 방법을 설계하여 사용자 평가에 적용하였다.

사용자 평가는 크게 암시적 피드백 특성이 실제 사용자의 곡 선호도를 잘 반영하는지 평가하기 위한 ‘음악 선호도 모델 평가’ 단계와, 특성이 협업적 음악 추천 알고리즘의 입력으로 사용되었을 때 추천 결과에 차이가 있는지 평가하기 위한 ‘음악 추천 알고리즘에서의 성능 평가’ 단계로 나뉜다. 본 장에서는 먼저 전반적인 평가 과정에 대해 설명한 후, 세부 절을 통해 각 평가 단계의 설계에 대해 상세히 서술한다.

제 1 절 평가 과정



[그림 7] 전체 사용자 평가 과정

[그림 7]은 본 연구의 전체 평가 과정을 도식화한 것이며, 각 단계에 따른 설명은 아래와 같다.

- ① 사용자 평가를 수행할 참여자를 모집한다.
- ② 참여자는 사용자 평가 전 인구통계학적 정보 및 청취 행태와 관련된 설문을 수행하며, 이 때 사용자 평가를 위한 연구 데이터를 수집을 위해 자신의 Last.fm 아이디를 제공한다.
- ③ 참여자로부터 제공받은 Last.fm 아이디를 쿼리로 하여 Last.fm API를 사용해 참여자의 전체 음악 청취 기록을 수집한다.
- ④ 수집한 참여자의 청취 기록을 사용하여, 재생 횟수, 누적 가중 재생을 및 재생횟수-평균 가중 재생을 곱 총 3가지의 암시적 피드백 특성을 가공한다.
- ⑤ 가공된 암시적 피드백 특성들을 입력으로 하여, 암시적 피드백에 특화된 협업적 음악 추천 알고리즘인 로지스틱 행렬분해를 사용하여 추천 리스트 총 3개를 도출한다.

- ⑥ 참여자는 음악 선호도 모델에 대한 평가를 수행한다.
- ⑦ 참여자는 음악 추천 결과에 대한 평가를 수행한다. 평가를 완료한 참여자에게는 보상을 지급한다.
- ⑧ 사용자 평가 결과를 분석한다.

참여자들은 해외 온라인 커뮤니티를 통해 모집되었으며, 평가에 참가하기 위한 조건은 다음과 같이 제시되었다.

- Last.fm 서비스에 가입하여 최소 5년 이상 음악을 scrobbling⁵⁾ 해 왔어야 함
- 영어권 국가에 거주하거나, 청취하는 곡들 중 대부분이 영어권 음악이어야 함

협업적 추천 알고리즘은 크게 이웃-기반(neighborhood-based)과 모델-기반(model-based) 방식으로 나뉘는데, 이들은 모두 같은 아이টে에 대해 평점을 제공하는 사용자의 수가 많을수록 추천의 정확도 및 신뢰도가 높아진다는 공통점을 갖는다. 이러한 이유로 연구에 사용된 LFM-1b 데이터셋의 데이터 수집 기간인 2013년 1월부터 2014년 8월까지 Last.fm 서비스를 통해 음악을 청취한 기록이 존재해야 한다는 조건은 사용자 평가를 통해 제시되는 추천 결과의 신뢰도를 높이기 위해 필수적이다. 비슷한 근거로, 추천된 곡들의 언어적, 문화적 일관성을 높이기 위해 두 번째 조건을 추가하였다.

참여자 모집 공고를 통해 초반에 모집된 참여자는 총 16명이었으며, 이들은 18세 미만의 참여자 1명, 18-24세의 참여자 3명, 25-34세의 참여자 7명, 35-44세의 참여자 5명으로 구성되었다. 이들 중 남성이 14명, 여성이 2명였으며, 영어권 국가(미국, 캐나다, 영국, 호주)에 거주하는 참여자는 10명이

5) Last.fm 서비스와 연동되는 음악 플레이어를 통해 음악을 청취함으로써 Last.fm에 청취 기록을 남기는 것

었다. 참여자 중 다수(약 69%)가 Spotify 서비스를 사용하여 Last.fm에 scrobble 하고 있다고 답했으며, 이 외에 4명은 기타 음악 스트리밍 서비스(iTunes, Deezer, Google Play Music)를, 나머지 2명은 비스트리밍 방식으로 음악을 청취한다고 답하였다. 전체 참여자 중 14명이 사용자 평가의 모든 과정을 완료했으며, 1명은 청취 행태 설문에만 참여, 1명은 청취 행태 설문 및 사용자 평가 중 음악 선호도 모델 평가 과정까지만 참여하였다. 사용자 평가 과정을 모두 완료한 참여자들에게는 \$10 상당의 보상이 지급되었으며, 중간까지 참여한 참여자에게는 부분적으로 보상을 지급하였다.

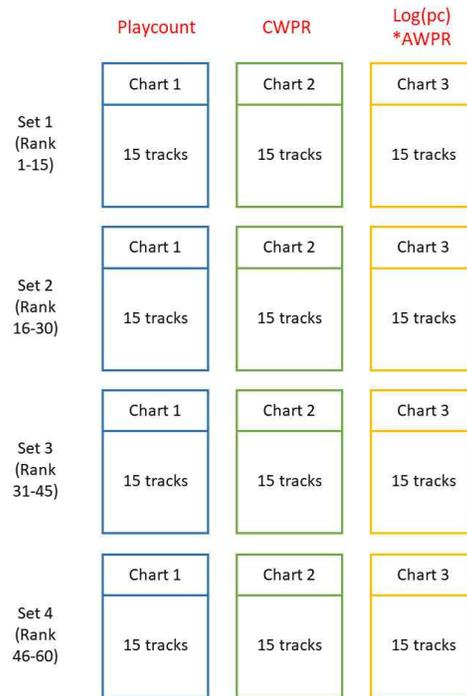
제 2 절 음악 선호도 모델 평가

암시적 피드백 특성 자체를 평가하기 위한 가장 직접적인 방법은 특성이 사용자들의 곡에 대한 선호도를 실제로 잘 나타내는지 사용자들에게 평가하도록 하는 것이다. 이를 실현할 수 있는 방법으로 고려해볼 수 있는 것은 현재 상용되고 있는 음악 서비스들의 일부에서 채택하고 있는 방식과 유사하게, 사용자가 청취한 곡들을 암시적 피드백 특성을 기준으로 정렬한 후 특정 순위에 있는 곡들을 사용자들에게 보여주는 것이다. 본 연구에서는 두 가지 암시적 피드백 특성을 제안하였으며, 이를 기존 음악 추천 시스템에서 일반적으로 사용되고 있는 재생횟수와 비교하는 것을 목적으로 하였다. 이에 따라 총 세 가지의 암시적 피드백 특성을 사용하여 사용자들이 청취한 곡들을 정렬한 후, 정렬된 곡 리스트를 기반으로 두 가지 방법을 사용하여 암시적 피드백 특성의 곡 선호도 모델로써의 정확도를 평가하고자 하였다. 각 방법에 대한 상세 설명은 다음 세부 절에서 제시한다.

4.2.1. 평가 문항 1

첫 번째 평가 문항에서는 참여자들에게 재생횟수, 누적 가중 재생을 및 재생횟수-평균 가중 재생을 곱의 세 가지 암시적 피드백 특성을 사용하여 정렬한 곡 순위 리스트의 일부를 제시하고, 각 리스트가 나타내는 곡들의 순위가 자신이 실제 해당 곡들에 대해 생각하는 순위와 비교하여 얼

마나 유사한지를 10점 척도로 평가하게 하였다. 이 때, 하위 순위로 갈수록 사용자들의 점수에 차이가 있는지 확인하기 위해 각 암시적 피드백 특성을 사용하여 정렬된 1-60순위까지의 트랙 리스트 세 개를 15순위씩 4세트로 나누어 제시하였다. [그림 8]은 첫 번째 문항에서 참여자들에게 주어진 순위 리스트들의 구성도이며, [그림 9]은 임의 참여자에게 주어진 첫 번째 문항의 실제 모습의 예시이다.



[그림 8] 음악 선호도 모델 평가 첫 번째 문항 구성도

A. How similar are these rankings to your real preference rankings for these tracks? Please rate each rank list between 1-10.

(Top 1-15)

Chart 1		Chart 2		Chart 3	
1	Ulvetid	Kvelertak	1	Ulvetid	Kvelertak
2	25 Or 6 To 4	Chicago	2	Sky Ride	Tweak Bird
3	Sky Ride	Tweak Bird	3	Mind Eraser, No Chaser	Them Crooked Vultures
4	Gloria Lewis	Kyuss	4	Mission from God	The Eighties Matchbox B-Line Disaster
5	Mind Eraser, No Chaser	Them Crooked Vultures	5	Riptin	The Eighties Matchbox B-Line Disaster
6	Mission from God	The Eighties Matchbox B-Line Disaster	6	Gloria Lewis	Kyuss
7	Riptin	The Eighties Matchbox B-Line Disaster	7	Round Trippin'	Tweak Bird
8	Tangy Zizzle	Kyuss	8	Tangy Zizzle	Kyuss
9	Round Trippin'	Tweak Bird	9	Man for All Seasons	The Eighties Matchbox B-Line Disaster
10	Man for All Seasons	The Eighties Matchbox B-Line Disaster	10	Lights In Lines	Tweak Bird
11	Lights In Lines	Tweak Bird	11	New Fang	Them Crooked Vultures
12	Shine	Kyuss	12	Beyond	Tweak Bird
13	New Fang	Them Crooked Vultures	13	Desire Be Desire Go	Tame Impala
14	Beyond	Tweak Bird	14	Shivers	Tweak Bird
15	Reptiles	Them Crooked Vultures	15	No One Loves Me & Neither Do I	Them Crooked Vultures
List Score(1 to 10):		<input type="text" value="1"/>	List Score(1 to 10):		<input type="text" value="1"/>
11		Reptiles	11		New Fang
12		You Think I Ain't Worth A Dollar, But I Feel Like A Millionaire	12		Desert Sessions
13		No One Loves Me & Neither Do I	13		Them Crooked Vultures

[그림 9] 음악 선호도 모델 평가 첫 번째 문항 예시

4.2.2. 평가 문항 2

재생 횟수는 3장에서 설명한 바와 같이, 사용자가 청취한 독립적인 곡의 수가 많을수록, 선호도 선상에서 하위 순위로 갈수록 같은 값을 가지는 곡들이 많아지는 현상을 보인다. 음악 재생율의 개념을 도입하게 되면 사용자가 청취한 곡들을 더 작은 단위로 표현하는 것이기 때문에 같은 값을 가지는 곡들이 생길 가능성이 크게 낮아지나, 여러 곡을 같은 선호도라고 가정하는 것과 비교하여 새로운 단위가 더 좋은 선호도 모델이라고 판단할 수 있는지는 별개의 문제이다.

이에 따라, 두 번째 문항에서는 참여자들이 청취한 곡들을 재생 횟수 기반으로 정렬했을 때 100-400위 사이에 위치한 곡들 중 재생 횟수가 같은 곡들 10개씩 2세트를 선정 한 후, 이들을 참여자에게 제시하여 각 세트의 곡들에 대해 선호하는 순위를 매기도록 지시하였다. 이러한 방식으로 곡들에 대해 사용자가 매긴 순위 리스트는 해당 곡들을 재생횟수, 누적 가중 재생율 및 재생횟수-평균 가중 재생율 곱 특성 각각을 기반으로 정렬하여 도출된 순위 리스트와 비교할 수 있게 되며, 사용자가 제공한 곡 순위 리스트와 유사한 순위를 도출한 특성일수록 사용자의 곡 선호도를 유사하게 표현한다고 할 수 있다. [그림 10]은 임의 참여자에게 주어진 두 번째 문항의 실제 모습의 예시이다.

B. Followings are lists of some tracks that you listened to. Please decide the ranking for these tracks based on your overall preference for them.

****All tracks in a list must have independent rank values. ****

Track list 1			Track list 2		
Pale Blue Eyes	The Velvet Underground	Rank 1 ▾	Freedom Run	Kyuss	Rank 1 ▾
Phototropic	Kyuss	Rank 1 ▾	From Time to Time	Ride	Rank 1 ▾
Exoskeleton, Part II	Dozer	Rank 1 ▾	Ondulé	Mathieu Boogaerts	Rank 1 ▾
Ants In The Kitchen	Masters of Reality	Rank 1 ▾	Green Onions	Booker T. & the MG's	Rank 1 ▾
Mudfly	Kyuss	Rank 1 ▾	Manolete	Yawning Man	Rank 1 ▾
Molten Universe	Kyuss	Rank 1 ▾	Plant Your Seed	Brant Bjork	Rank 1 ▾
Sundown	Torche	Rank 1 ▾	Treehouse Song	Nina Nastasia	Rank 1 ▾
Doraldina's Prophecies	Masters of Reality	Rank 1 ▾	Immigrant Song	Led Zeppelin	Rank 1 ▾
Reservations	Tweak Bird	Rank 1 ▾	Mind Over All	Hatebreed	Rank 1 ▾
Tilt-A-Whirl	Masters of Reality	Rank 1 ▾	Destroy Everything	Hatebreed	Rank 1 ▾

[그림 10] 음악 선호도 모델 평가 두 번째 문항 예시

제 3 절 음악 추천 알고리즘에서의 성능 평가

곡에 대한 암시적 피드백은 근본적으로는 사용자들의 선호도를 모델링하는 데에 있으며, 이를 위한 평가는 2절에서 설명한 음악 선호도 모델

평가 과정을 통해 수행하였다. 다음으로 암시적 피드백의 궁극적 목적은 추천의 성능을 높이는 것에 있으며, 이에 따라 3절에서는 기존 및 제안한 암시적 피드백 특성이 음악 추천 알고리즘에 적용되었을 때의 추천 결과에 대한 평가를 수행하는 과정을 설명한다.

4.3.1. 로지스틱 행렬 분해 알고리즘

본 연구에서 제안한 암시적 피드백 특성의 추천 알고리즘에서의 성능을 평가하기 위해, 사용자의 아이템에 대한 피드백 혹은 인터랙션 데이터가 핵심적인 부분을 차지하는 협업적 필터링 방식 중 행렬분해 알고리즘을 사용하였다. 추천 대상 사용자의 아이템 벡터와 유사도가 높은 벡터를 가진 상위 일부 사용자들을 이웃으로 가정하여 이들의 인터랙션 데이터만을 기반으로 아이템에 대한 평점을 추정하는 이웃-기반(neighborhood-based) 방식과 달리, 행렬분해 알고리즘에서는 사용자-아이템 행렬 중 사용자의 소비 이력이 없는 모든 아이템들에 대한 평점을 추정한 후 평점이 높은 아이템 순으로 대상 사용자에게 추천이 이루어진다.

그 중 확률-기반의 행렬분해 알고리즘에서는 사용자의 특정 아이템에 대한 추정된 인터랙션 값을 사용자 잠재 벡터와 아이템 잠재 벡터 간의 내적값이 아닌 확률 모델로 정의한다. 비교적 최근 공개된 로지스틱 행렬분해(Logistic Matrix Factorization) 알고리즘(Johnson, 2014)의 경우 이러한 확률 모델을 로지스틱 함수로 정의하고, 암시적 피드백 데이터를 사용한 행렬분해 방식 기반 협업적 필터링 추천 알고리즘의 가장 대표적인 연구(Hu et al., 2008)에서 제안한 방식을 차용하여 모델이 부정적인 암시적 피드백 값들의 영향을 제어할 수 있도록 확장하였다. 이러한 로지스틱 행렬분해 알고리즘이 기존의 암시적 피드백을 사용한 행렬분해 알고리즘의 성능을 능가하는 것이 검증되었음에 따라, 본 연구에서는 사용자 평가를 위한 음악 추천 알고리즘으로 로지스틱 행렬분해 알고리즘을 선정하였다.

4.3.2. 데이터 샘플링

앞서 설명한 바와 같이 행렬 분해 알고리즘을 사용하면 사용자-아이템 행렬에서 인터랙션이 존재하지 않는 사용자-아이템 쌍들에 대해 인터랙션 값을 추정하여 행렬을 채울 수 있게 된다. LFM-1b 데이터셋에는 약 12만 명의 사용자 데이터가 기록되어 있으나, 본 연구에서는 컴퓨팅 성능 및 시간의 한계로 인하여 사용자 평가를 위한 조건을 다수 선정하여 이를 기반으로 사용자 데이터를 일부 샘플링하였다.

먼저 추천 결과의 언어적, 문화적 일관성을 위해, 평가 참여자 모집 시의 조건과 유사하게 LFM-1b 데이터셋의 사용자 프로필 데이터의 국적을 영어권 국가(미국, 캐나다, 영국, 호주)로 표기한 사용자들 중 3000명을 임의로 샘플링하였다. 이 후 추가적으로, 추천 알고리즘에 사용되는 사용자들의 청취 데이터의 신뢰도를 높이기 위한 목적으로 청취한 독립적인 트랙의 수가 500곡 미만, 청취 이벤트의 수가 1000개 미만인 사용자들을 필터링하였다. 이에 따라 로지스틱 행렬 분해 알고리즘에 사용된 사용자 데이터의 수는 실제 평가 참여자를 제외하고 1595개였으며, 샘플링된 데이터의 상세 통계는 [표 4]와 같다.

데이터 구분	수치
사용자 수	1,595
전체 청취 이벤트 수	13,510,310
전체 독립 트랙 수	4,402,995
청취 이벤트 중앙값	2,288.0
독립 트랙 수 중앙값	6,569.0

[표 4] 샘플링된 LFM-1b 사용자 데이터 상세 통계

4.3.3. 음악 추천 리스트 평가 방법

본 연구에서는 사용자 평가 방법을 적용했기 때문에 오프라인 평가 방법과 달리 추천 결과에 대해 실제 참여자들에게 다양한 기준을 사용하여 평가할 것을 요청했으며, 추천된 곡들에 대해 평가 척도 중 관련성(relevance), 우연성(serendipity), 새로움(novelty)을 기준으로 제시하였다. 이에 따라 세 번째 문항에서는 참여자 각각에게 재생 횟수, 누적 가중 재생율, 재생횟수-평균 가중 재생율 곱 특성 각각을 암시적 피드백으로 사용했을 때 도출된 추천 리스트 총 세 가지가 주어졌으며, 각 리스트는 10개 곡으로 구성되었다. 참여자는 먼저 리스트 내의 각 곡에 대해 세 가지 척도를 기준으로 평가하였으며, 이 후 추가적으로 각 추천 리스트에 대한 만족도 점수를 제공하였다. 이 때 관련성은 10점 척도, 우연성 및 새로움은 ‘그렇다/아니다’의 바이너리(binary) 척도로 평가되었다. [그림 11]은 임의 참여자에게 주어진 추천 결과 평가 문항의 실제 모습의 예시이다.

Followings are 3 lists of song recommendations, each created using different preference models. Please listen to the songs using your music player, and evaluate each of them based on the 3 criteria. When you are done evaluating all songs in a list, give an overall score to the whole list.

- **Relevance:** How relevant is this recommended song to your music taste, in light of your music listening history? Or how much sense does it make?
- **Novelty:** Have you ever heard this song, either intentionally or unintentionally?
- **Serendipity:** Was this recommendation both surprising(unexpected), and pleasant at the same time?

Recommendation List 1					
Rank	Title	Artist	Relevance	Novelty	Serendipity
1	Count In Fives	The Horrors	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
2	Last Night I Dreamt That Somebody Loved Me	The Smiths	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
3	Just a Girl	No Doubt	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
4	Sunday Morning	The Velvet Underground	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
5	Tomorrow Never Knows	The Beatles	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
6	Decades	Joy Division	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
7	Do the Evolution	Pearl Jam	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
8	Dissolve Me	Alt-J	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
9	Show of Strength	Echo & the Bunnymen	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
10	Temecula Sunrise	Dirty Projectors	1 ▾	<input type="radio"/> yes <input checked="" type="radio"/> no	<input type="radio"/> yes <input checked="" type="radio"/> no
List Score(1 to 10):			1 ▾		

[그림 11] 음악 추천 결과 평가 문항 예시

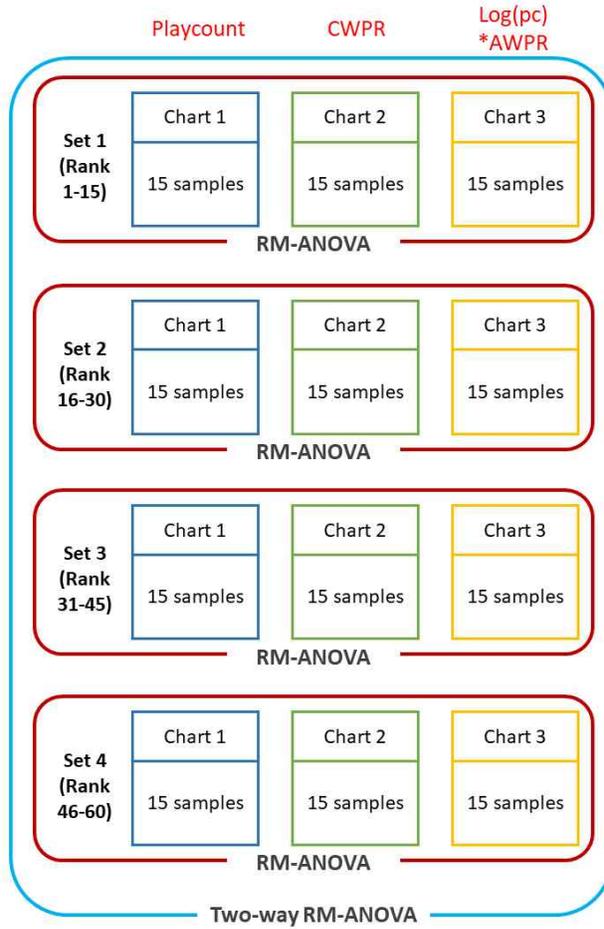
제 5 장 연구 결과

본 장에서는 먼저 사용자 평가를 수행하여 얻은 참여자들의 데이터를 통계적으로 분석하여 얻은 각 평가 문항에 관한 결과를 설명한다. 이 후 사용자 평가를 분석하여 얻은 결과를 정리하고, 재생율 개념의 도입으로 인한 추천 성능 개선의 가능성, 연구 설계의 한계점, 음악 선호도 모델 및 추천 시스템과의 상관관계 등에 관해 고찰한다.

제 1 절 음악 선호도 모델 평가 결과

5.1.1. 평가 문항 1 결과

첫 번째 문항에 대한 결과는 크게 두 가지 방법으로 분석되었다. 첫 번째 문항에서 사용자들은 총 네 세트의 순위 범위(1-15, 16-30, 31-45, 46-60) 리스트 각각에 대해 10점 척도의 점수를 제공하였다. 이에 따라 먼저 각 순위 범위 리스트의 점수 간에 차이가 있는지 보고자 각 순위 범위에 대해 세 종류의 암시적 피드백 특성을 독립 변수, 리스트에 대한 참여자들의 점수를 종속 변수로 하여 반복측정 분산분석(Repeated Measures ANOVA)을 수행하였다. 다음으로는 전체 순위 범위(1-60)에 대한 리스트 점수를 취합하여, 순위 범위와 암시적 피드백 특성 모두를 독립 변수로 하여 이원 반복측정 분산분석(Two-way Repeated Measures ANOVA)을 수행하였다. [그림 12]는 음악 선호도 모델 평가의 첫 번째 문항 결과에 대해 통계 분석을 수행한 구조를 나타낸 것이다.



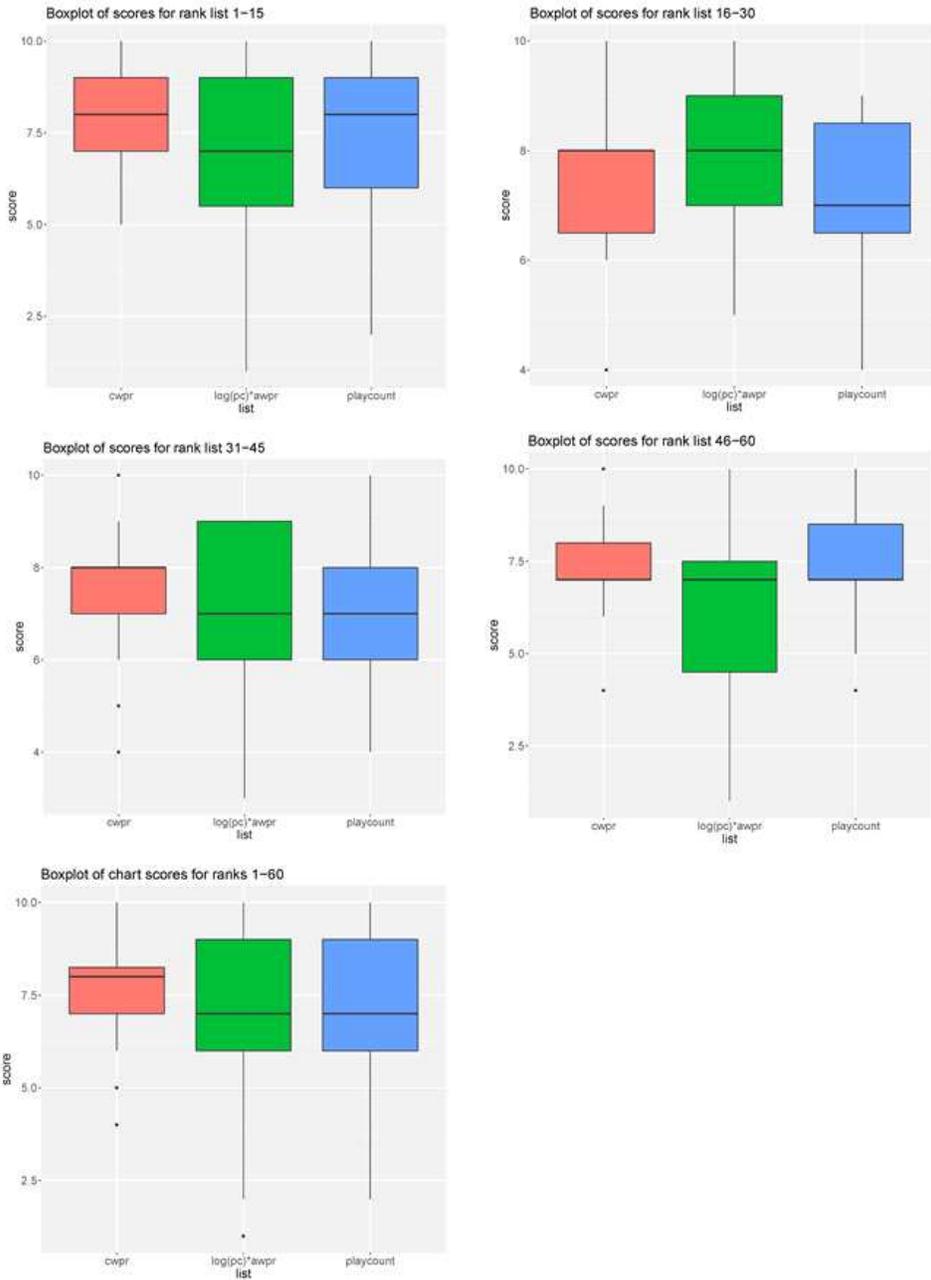
[그림 12] 첫 번째 문항의 결과에 대해 통계 분석을 수행한 구조

[표 5]는 첫 번째 문항에 대해 각 차트의 순위 범위별 평균 및 표준편차와, 위와 같은 방법으로 통계 분석을 수행한 결과를 나타내며, [그림 13]은 개별 순위 범위 및 전체 순위에 대해 수집된 점수의 상자그림(boxplot)을 나타낸다.

순위 범위	차트 1 (재생횟수)		차트 2 (누적 가중 재생율)		차트 3 (재생횟수-평균 가중 재생율 곱)		RM-ANOVA 결과 (p-value)
	평균	표준편차	평균	표준편차	평균	표준편차	
1-15	7.33	2.26	<u>7.87</u>	1.46	6.87	2.39	0.272
16-30	7.13	1.6	<u>7.27</u>	1.71	<u>7.67</u>	1.4	0.470
31-45	7	1.78	<u>7.33</u>	1.5	7	1.96	0.703
46-60	<u>7.33</u>	1.76	7.13	1.68	6.27	2.87	0.076
1-60	7.22	1.82	<u>7.43</u>	1.59	6.98	2.24	(순위) 0.72 (특성) 0.30 (특성:순위) 0.25

* 1-60 순위 범위의 경우 이원 반복측정 분산분석 결과에 대한 p-value를 나타냄

[표 5] 첫 번째 문항의 각 차트의 순위 범위별 점수 평균 및 표준편차와
통계 분석 결과



[그림 13] 개별 및 전체 순위 범위에 대해 수집된 점수의 상자그림(boxplot)

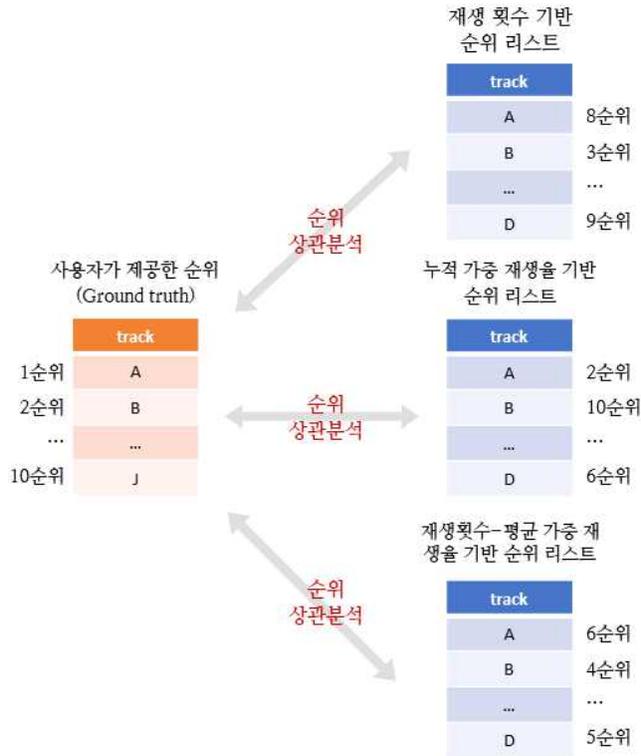
결과에 의하면 각 차트의 점수 차이는 통계적으로는 유의하지 않았으나, 46-60 순위 범위의 점수들 간에는 거의 유의한 수준의 차이가 있어 ($p=0.076$) 추가적으로 대응표본 t-검정(paired t-test)을 수행하였다.

전체 순위 범위(1-60)에 대해 이원 반복측정 분산분석을 수행했을 때, 각 암시적 피드백 특성에 따른 차트 점수간의 차이는 유의하지 않았으며 ($p=0.3$), 순위 범위는 차트 점수에 영향을 미치지 않는 것으로 나타났다 ($p=0.72$). 또한 순위 범위와 특성 간의 상호작용은 없었다($p=0.25$).

첫 번째 문항에 대한 통계 분석 결과 통계적으로 유의한 차이는 없었으나, [표 5]에서 확인할 수 있듯이 전반적으로 누적 가중 재생율을 사용하여 정렬된 차트에 대한 참여자들의 평균 점수가 가장 높았으며, 다음으로는 재생 횟수, 재생횟수-평균 가중 재생율 특성 순으로 점수가 높았다.

5.1.2. 평가 문항 2 결과

두 번째 문항을 통해 15명의 참여자 개인별로 곡 20개에 대한 1-10까지의 순위 리스트가 2개씩 수집되었다. 앞서 4장에서 설명한 바와 같이, 참여자가 제공한 순위 리스트를 각 암시적 피드백 특성을 기반으로 해당 곡들을 정렬하여 생성된 순위 리스트와 비교함으로써 순위 리스트 간의 상관관계가 있는지 보고자 하였다. [그림 14]는 음악 선호도 모델 평가의 두 번째 문항 결과에 대해 통계 분석을 수행한 구조를 나타낸 것이다.



[그림 14] 두 번째 문항의 결과에 대해 통계 분석을 수행한 구조

순위 상관관계 분석에는 Kendall's τ (타우) 상관계수가 사용되었으며, 15명의 참여자가 제공한 순위 리스트 총 30개(15×2) 각각에 대해 세 가지 암시적 피드백 특성이 생성한 순위 리스트 각각과의 순위 상관분석이 이루어졌다. 이에 따라 각 암시적 피드백 특성에 대해 30개씩의 상관계수가 도출되었다. 참여자가 순위를 매기는 작업은 각 순위 리스트마다 독립적이라고 판단되었음에 따라, Fisher의 최소유의차이(Least Significant Distance) 방법을 사용하여 각 특성별 30개의 상관계수에 대한 각각의 신뢰 구간을 취합하여 하나의 신뢰 구간을 구하였다. 이 때, 각 특성에 대해 상관계수의 부호가 양수 및 음수 모두 관측되었기 때문에 추가적으로 부호가 같은 결과끼리 취합하여 각 부호(양의 상관관계/음의 상관관계)에 대한 신뢰구간을 각각 구하였다. [표 6]은 각 특성에 대해 Fisher의 최소유의차이 방법을 수행하여 도출된 분석 결과를 나타낸다. 표의 모든 수치는 유의확률(p-value)을

나타낸다.

	재생 횟수	누적 가중 재생율	재생 횟수-평균 가중 재생율 곱
$\tau > 0$	0.831	0.325	0.448
$\tau < 0$	0.418	0.312	0.168
전체 τ	0.729	0.276	0.217

[표 6] 각 암시적 피드백 특성 기준 순위 리스트의 순위 상관분석 결과

순위 상관분석 결과, 참여자가 제공한 곡 순위 리스트와 유의한 상관 관계를 보이는 순위를 생성할 수 있는 암시적 피드백 특성은 없는 것으로 나타났다. 또한, 양의 상관계수에 대한 유의확률보다 음의 상관계수에 대한 유의확률이 대체적으로 낮게 나타나는 것을 볼 수 있다. 본 문항의 경우, 이와 같은 분석 결과를 통해 순위 선정을 위해 참여자들에게 제시된 곡들의 선정 방법에 있어서 여러 문제가 있을 수 있음을 인지하게 되었으며, 이에 대한 내용은 3절에서 종합적으로 서술한다.

제 2 절 음악 추천 알고리즘에서의 성능 평가 결과

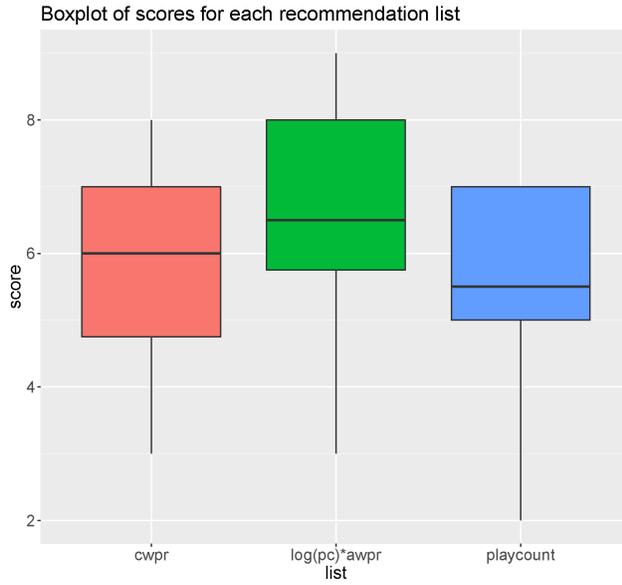
사용자 평가의 마지막 문항에서는 재생 횟수, 누적 가중 재생율 및 재생 횟수-평균 가중 재생율 곱 특성 세 가지를 협업적 추천 알고리즘에 적용하여 도출된 곡 추천 리스트 각각에 대해, 리스트 내의 곡 각각 및 리스트 전체에 대한 참여자들의 평가 점수가 수집되었다. 이에 따라 본 절에서는 리스트 전체에 대한 평가와 개별 곡에 대한 평가를 통계적으로 분석한 결과를 세부 절로 나누어 상세히 설명한다.

5.2.1. 추천 곡 리스트 평가 결과

음악 추천 알고리즘에서의 암시적 피드백 특성들의 성능을 평가하기 위한 마지막 문항에는 14명의 참여자들이 참여하였으나, 2명의 참여자가 추천된 개별 곡에 대한 점수는 제공하고 전체 리스트에 대한 점수는 제공하지 않아 이들을 제외한 12명의 참여자에 대한 평가 결과를 분석하였다. [표 7]은 각 추천 곡 리스트에 대해 참여자들이 제공한 전체 점수의 평균 및 표준편차를 나타내며, [그림 15]는 각 추천 곡 리스트별 전체 점수의 상자그림(boxplot)을 나타낸다. 점수 분포를 통해 재생 횟수-평균 가중 재생율 곱 특성을 추천 알고리즘에 적용했을 때 도출된 추천 곡 리스트에 대한 참여자들의 점수가 가장 높은 경향을 보였으며, 재생 횟수의 경우 가장 낮은 평균 점수를 나타냈다.

재생 횟수 추천 곡 리스트		누적 가중 재생율 추천 곡 리스트		재생 횟수-평균 가중 재생율 곱 추천 곡 리스트	
평균	표준편차	평균	표준편차	평균	표준편차
5.5	1.51	5.92	1.62	<u>6.42</u>	1.78

[표 7] 각 추천 곡 리스트에 대한 전체 점수의 평균 및 표준편차



[그림 15] 각 추천 곡 리스트별 전체 점수의 상자그림(boxplot)

이 후 각 특성별로 도출된 추천 곡 리스트에 대한 전체 점수 간에 차이가 있는지 보고자 하였으며, 이에 따라 세 종류의 암시적 피드백 특성을 독립 변수, 추천 곡 리스트에 대한 참여자들의 점수를 종속 변수로 하여 반복측정 분산분석(Repeated Measures ANOVA)을 수행하였다. 그 결과 통계적으로는 전체 점수간의 차이가 유의하지 않은 것으로 나타났다($F_{2,22}=1.152$, $p=0.334$). 추후 실험 설계의 보완을 통한 성능 개선의 여지가 있는지 확인하기 위해 대응표본 t-검정(paired t-test)를 수행하였으며, 그 결과는 [표 8]와 같다. 아래의 결과를 통해 재생 횟수와 재생 횟수-평균 가중 재생을 곱 특성 간의 점수 차이에 대한 유의확률이 작은 축에 속하는 것을 알 수 있다 ($p = 0.19$).

재생 횟수 & 누적 가중 재생율		누적 가중 재생율 & 재생 횟수-평균 가중 재생율 곱		재생 횟수 & 재생 횟수-평균 가중 재생율 곱	
t	p-value	t	p-value	t	p-value
-0.658	0.524	-0.971	0.3524	-2.362	0.190

[표 8] 추천 곡 리스트 전체 점수의 대응표본 t-검정 수행 결과

5.2.2. 개별 추천 곡 평가 결과

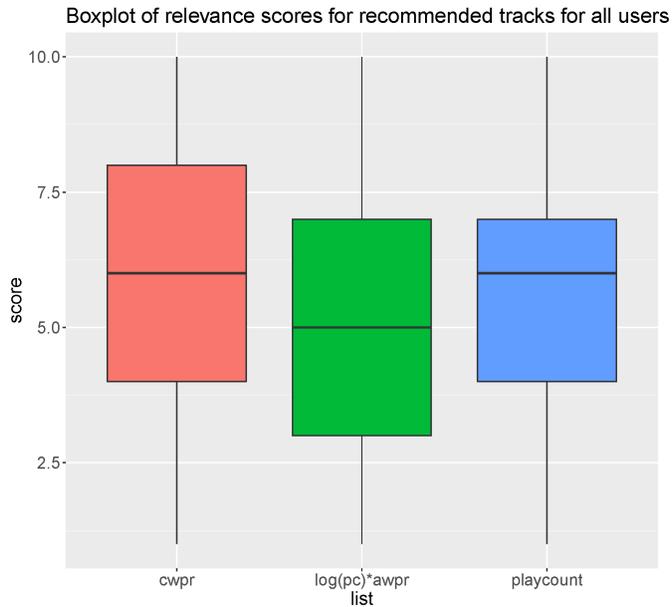
주어진 추천 곡 리스트 내의 각 곡에 대해 세 가지 척도(관련성, 새로움, 유연성)에 대한 총 14명의 참여자의 평가 점수가 수집되었다. 수집된 평가 데이터는 평가 척도별 해석을 위해, 각 척도에 대해 개별적으로 통계 분석이 이루어졌다.

관련성(Relevance)

먼저 각 암시적 피드백 특성별로 도출된 세 가지 추천 곡 리스트 내의 곡들에 대해 참여자들이 평가한 관련성 점수의 평균 및 표준편차는 [표 9]에 나타나 있으며, 이 데이터의 상자그림(boxplot)은 [그림 16]과 같다. 이 때 참여자 14명이 각 추천 곡 리스트 내의 곡 10개씩에 대한 관련성 점수를 제공했으므로, 각 리스트의 평균 관련성 점수에 사용된 샘플 수는 140이다.

누적 가중 재생을 추천 곡 리스트		재생 횟수-평균 가중 재생을 곱 추천 곡 리스트		재생 횟수 추천 곡 리스트	
평균	표준편차	평균	표준편차	평균	표준편차
<u>6.09</u>	2.54	5.41	2.72	5.5	2.53

[표 9] 각 추천 곡 리스트에 대한 관련성 점수의 평균 및 표준편차



[그림 16] 각 추천 곡 리스트별 관련성 점수의 상자그림(boxplot)

점수 분포를 통해 누적 가중 재생을 암시적 피드백으로 하여 추천된 곡들에 대한 관련성 점수가 가장 높으며, 재생 횟수-평균 가중 재생을 곱을 사용하여 추천된 곡들에 대한 관련성 점수가 가장 낮은 경향을 보이는 것을 알 수 있다.

다음으로 추천 곡 리스트 간의 관련성 점수에 유의미한 차이가 있는지 확인해야 하나, 한 명의 참여자가 서로 다른 곡 10개에 대해 제공한 관련성 점수에 대한 상호 독립성을 판단하는 것이 모호하여 같은 특성을 사용하여 추천된 10곡에 대한 관련성 점수들의 평균값을 분석에 사용하였다. 이에 따라 세 종류의 암시적 피드백 특성을 독립 변수, 추천 곡 리스트에 대한 평균 관련성 점수를 종속 변수로 하여 반복측정 분산분석(Repeated Measures ANOVA)을 수행하였다.

분석 결과, 추천된 곡 리스트의 관련성 점수가 암시적 피드백 특성에 따라 차이가 있을 확률이 높은 것으로 나타났으며($F_{2,26}=2.69, p=0.086$), 이에 따라 사후검증을 위해 대응표본 t-검정(paired t-test)를 수행하였다. 사후검증 결과는 [표 10]과 같으며, 아래의 결과를 통해 누적 가중 재생율 특성이 재생 횟수-평균 가중 재생율 곱 특성에 비해 유의한 차이로 관련성 높은 곡들을 추천했으며($p=0.053$), 재생 횟수 특성에 비해서도 실험 설계의 보완을 통해 유의한 차이를 보일 가능성이 있음을 알 수 있었다($p=0.14$).

재생 횟수 & 누적 가중 재생율		누적 가중 재생율 & 재생 횟수-평균 가중 재생율 곱		재생 횟수 & 재생 횟수-평균 가중 재생율 곱	
t	p-value	t	p-value	t	p-value
-1.571	0.140	2.1259	0.053	0.376	0.713

[표 10] 추천 곡 리스트에 대한 관련성 점수의 대응표본 t-검정 수행 결과

새로움(Novelty)

각 암시적 피드백 특성별로 도출된 세 가지 추천 곡 리스트 내의 곡들에 대해 참여자들이 평가한 새로움 점수는 ‘그렇다/아니다’의 바이너리 값으로 수집되었으며, 통계적인 분석을 위해 각 리스트에 대한 새로움 점수는 수식

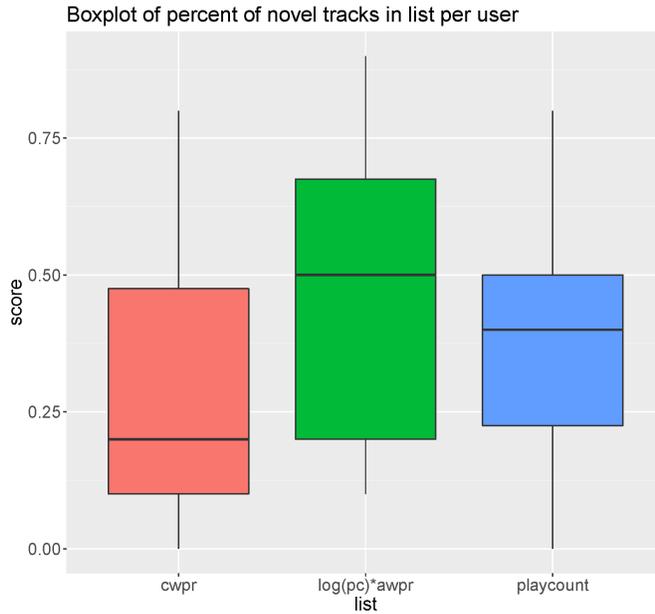
(1)과 같이 리스트 내 전체 곡의 수(10) 중 참여자가 ‘그렇다(새롭다)’라고 평가한 곡의 수의 비율로 정의하였다.

$$\text{새로움 (Novelty)} = \frac{\text{'그렇다(새롭다)'라고 평가한 곡의 수}}{\text{리스트 내 전체 곡 수}} \quad (1)$$

각 암시적 피드백 특성별로 도출된 세 가지 추천 곡 리스트에 대해 참여자들이 평가한 새로움 점수의 평균 및 표준편차는 [표 11]에 나타나 있으며, 이 데이터의 상자그림(boxplot)은 [그림 17]과 같다.

누적 가중 재생을 추천 곡 리스트		재생 횟수-평균 가중 재생을 곱 추천 곡 리스트		재생 횟수 추천 곡 리스트	
평균	표준편차	평균	표준편차	평균	표준편차
0.293	0.23	0.45	0.27	0.371	0.23

[표 11] 각 추천 곡 리스트에 대한 새로움(novelty) 점수의 평균 및 표준편차



[그림 17] 각 추천 곡 리스트별 새로움(Novelty) 점수의 상자그림(boxplot)

점수 분포를 통해 재생 횟수-평균 가중 재생을 곱 특성을 암시적 피드백으로 하여 추천된 곡들에 대한 새로움 점수가 가장 높으며, 누적 가중 재생을 특성을 사용하여 추천된 곡들에 대한 새로움 점수가 가장 낮은 경향을 보이는 것을 알 수 있다.

추천 곡 리스트 간의 새로움 점수에 유의미한 차이가 있는지 확인하기 위해, 세 종류의 암시적 피드백 특성을 독립 변수, 추천 곡 리스트에 대한 새로움 점수를 종속 변수로 하여 반복측정 분산분석(Repeated Measures ANOVA)을 수행하였으며, 그 결과 추천된 곡 리스트의 새로움 점수가 암시적 피드백 특성에 따라 차이가 있을 확률이 높은 것으로 나타났다($F_{2,26}=2.356$, $p=0.115$), 이에 따라 사후검증을 위해 추천 곡 리스트 간 대응표본 t-검정(paired t-test)를 수행하였다. 사후검증 결과는 [표 12]와 같으며, 아래의 결과를 통해 재생 횟수-평균 가중 재생을 곱 특성이 누적 가중 재생 및 재생 횟수 특성에 비해 비교적 새로운 곡들을 추천했을 가능성이 높은 편임을 알 수 있다($p=0.092$, $p=0.181$).

재생 횟수 & 누적 가중 재생율		누적 가중 재생율 & 재생 횟수-평균 가중 재생율 곱		재생 횟수 & 재생 횟수-평균 가중 재생율 곱	
t	p-value	t	p-value	t	p-value
1.091	0.295	-1.822	0.092	-1.412	0.181

[표 12] 추천 곡 리스트에 대한 새로움 점수의 대응표본 t-검정 수행 결과

우연성(Serendipity)

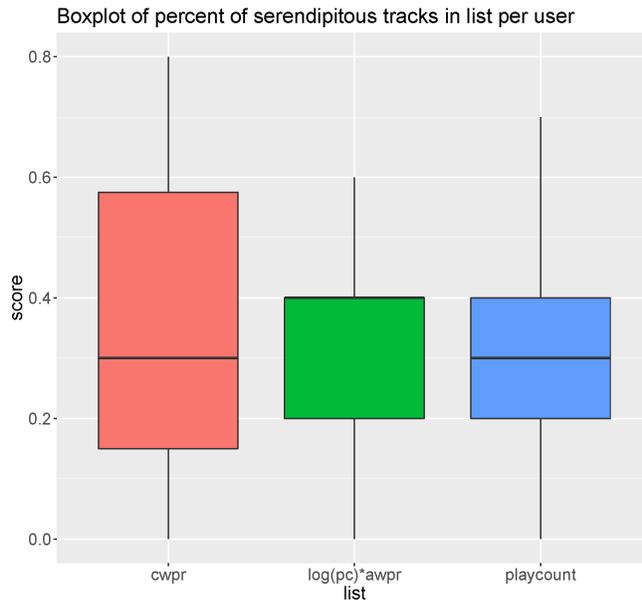
우연성 척도 역시 바이너리 값으로 수집되었으며, 이에 따라 우연성 척도는 수식 (2)와 같이 정의할 수 있다.

$$\text{우연성 (Serendipity)} = \frac{\text{'그렇다(우연이다)'라고 평가한 곡의 수}}{\text{리스트내 전체 곡 수}} \quad (2)$$

각 암시적 피드백 특성별로 도출된 세 가지 추천 곡 리스트에 대해 참여자들이 평가한 우연성 점수의 평균 및 표준편차는 [표 13]에 나타나 있으며, 이 데이터의 상자그림(boxplot)은 [그림 18]과 같다.

누적 가중 재생율 추천 곡 리스트		재생 횟수-평균 가중 재생율 곱 추천 곡 리스트		재생 횟수 추천 곡 리스트	
평균	표준편차	평균	표준편차	평균	표준편차
<u>0.364</u>	0.27	0.321	0.2	0.307	0.19

[표 13] 각 추천 곡 리스트에 대한 우연성(serendipity) 점수의 평균 및 표준편차



[그림 18] 각 추천 곡 리스트별 우연성(Serendipity) 점수의 상자그림(boxplot)

점수 분포를 통해 누적 가중 재생을 곱 특성을 암시적 피드백으로 하여 추천된 곡들에 대한 우연성 점수가 가장 높으며, 나머지 두 특성은 큰 차이를 보이지 않는 것을 알 수 있다.

추천 곡 리스트 간의 우연성 점수에 유의미한 차이가 있는지 확인하기 위해, 세 종류의 암시적 피드백 특성을 독립 변수, 추천 곡 리스트에 대한 우연성 점수를 종속 변수로 하여 반복측정 분산분석(Repeated Measures ANOVA)을 수행하였다. 그 결과 추천된 곡 리스트의 우연성 점수가 암시적 피드백 특성에 따라 유의한 차이가 없는 것으로 나타났다($F_{2,26}=0.434$, $p=0.653$), 추천 곡 리스트 간 대응표본 t-검정(paired t-test)를 수행했을 때도 유의미한 우연성 점수 차이를 보이는 리스트 쌍은 없는 것으로 나타났다.

제 3 절 결과 정리 및 고찰

본 연구에서는 앞서 설명한 방식의 사용자 평가를 통해 총 세 가지의 암시적 피드백 특성이 사용자들의 곡 선호도를 잘 반영할 수 있는지와 이들이 협업적 추천 알고리즘에 적용되었을 때 추천 결과에 차이가 있는지를 검증하고자 하였다. 사용자 평가의 결과를 통계적으로 분석한 결과, 기존 음악 추천 시스템 분야에서 암시적 피드백으로 주로 사용되어 왔던 재생 횟수와, 본 연구를 통해 제안한 누적 가중 재생율과 재생 횟수-평균 가중 재생율 곱의 총 세 가지 암시적 피드백 특성 중 누적 가중 재생율을 사용하여 평가 참여자들이 청취한 곡의 순위를 매겼을 때 가장 높은 점수를 받는 경향을 보였다. 이는 특히 재생 횟수-평균 가중 재생율 곱 특성과 비교했을 때 차이가 더욱 컸으며($p=0.07$), 재생 횟수 특성과 비교해서도 실험의 보완을 통해 유의한 차이를 보일 수 있을 가능성을 시사했다($p=0.277$). 반대로 해석하면, 재생 횟수-평균 가중 재생율 곱 특성을 사용하여 사용자들의 곡 선호도를 모델링했을 때 가장 높은 점수를 받은 특성과 비교하여 거의 유의한 차이로 가장 낮은 점수를 받았음을 의미한다.

다음으로 각 암시적 피드백 특성을 실제 협업적 음악 추천 알고리즘에 적용하여 도출된 추천 결과에 대해 추천 곡 리스트 전체에 대한 만족도 점수와 리스트 내의 개별 곡들에 대해 관련성, 새로움 및 우연성을 기준으로 한 사용자들의 평가를 수집한 결과, 제안 특성 중 재생횟수-평균 가중 재생율 곱을 사용했을 때의 추천 곡 리스트 전체에 대한 만족도 점수가 기존의 재생 횟수에 비하여 높은 경향을 보였다($p=0.17$). 추천 리스트 내 개별 곡들에 대한 평가의 경우, 재생횟수-평균 가중 재생율 곱 특성이 누적 가중 재생율 특성과 비교하여 거의 유의한 차이로 새로운 곡들을 추천했으며($p=0.09$), 반대로 누적 가중 재생율 특성이 재생횟수-평균 가중 재생율 곱 특성에 비해 유의한 차이로 관련성이 높은 곡들을 추천하였다($p=0.05$). 다음 세부 절에서는 이와 같은 실험 결과를 바탕으로 고찰한 내용에 대해 서술한다.

5.3.1. 선호도 모델과 추천 결과와의 관련성

실험 결과에서 가장 주목할만한 점은 음악 선호도 모델로써의 성능과 음악 추천 알고리즘에서의 성능 평가를 수행했을 때의 재생횟수-평균가중 재생율 곱 특성의 상반되는 평가 결과이다. 이 특성은 선호도 모델로써는 가장 낮은 점수를 받았던 반면, 이 특성을 사용하여 도출된 곡 추천 리스트에 대한 만족도 점수는 가장 높았다. 이러한 결과만을 놓고 보면 ‘아이템에 대한 사용자들의 선호도를 잘 추정할수록 좋은 추천 성능으로 이어질 것이다’라는 가정은 틀린 것처럼 보인다. 그러나 본 연구에서 선호도 모델의 평가를 위해 설계한 사용자 평가 프레임워크는 이전에 시도된 바가 없었음에 따라 많은 한계점을 지니며, 이 때문에 연구 결과에도 해석의 여지가 많을 수 있다. 이에 따라 다음에서는 연구 설계의 한계와 평가 결과를 여러 관점에서 고찰한다.

먼저 사용자 평가의 첫 번째 문항에서는 참여자가 청취한 곡들을 세 가지 암시적 피드백 특성을 기반으로 정렬한 상위 60곡씩이 주어졌으며, 이러한 설계에 대한 이유는 사용자에게 평가의 부담을 과도하게 주지 않으면서, 사용자가 잘 인지하고 있는 곡들일수록 평가를 정확히 할 것이라고 판단되었기 때문이다. 그러나 곡 선호도 선상에서 가장 상위에 위치하는 곡들일수록 사용자가 이들 간의 순위를 정확히 매긴다는 것은 모호한 목표일 수 있으며, 이에 따라 첫 번째 문항에 대한 참여자들의 평가 결과가 유의한 차이를 보이지 않을 가능성이 높을 수 있다. 실제로, 추후 실험 보완의 용도로 참여자들로부터 받은 피드백 중에는 ‘세 차트 간에 큰 차이가 있지는 않다’, ‘상위 30곡은 서로 구별하기 힘들어서 결정하기 힘들었다’, ‘모든 리스트가 흥미로운 결과이기 때문에 선택하기 힘들다’ 등의 내용이 포함되었다.

선호도 선상의 상위 곡들에 대해서만이 아니라, 암시적 피드백 특성들의 중간 순위의 곡들에 대한 선호도 모델링 성능 역시 평가하기 위해 고안된 것이 사용자 평가의 두 번째 문항이다. 그러나 분석 결과 사용자가 곡들에 대

해 제공한 선호도 순위와 비교하여 상관관계가 존재하는 순위 리스트를 생성할 수 있는 특성이 없는 것으로 나타났으며, 결과를 통해 두 번째 문항에서 참여자들에게 제시된 곡들을 선정하는 기준에 문제가 있음을 인지하게 되었다. 재생 횟수를 기준으로 정렬했을 때 100-400위까지의 범위는 최상위 순위는 아닌 동시에, 사용자들이 충분히 해당 곡들을 기억하고 선호도를 판별할 수 있는 순위 범위인 것으로 판단된다. 그러나 재생횟수가 구분할 수 없는 곡들을 제안한 특성들이 제대로 구별할 수 있는지 확인하기 위해 참여자들에게 같은 재생횟수를 가지는 곡들을 제시한 것은 첫 번째 문항에서와 유사한 이유로, 선호도 선상에서 극도로 가까운 거리에 있는 곡들에 대한 평가라는 모호한 목표를 제시한 것과 같다. 이에 따라, 평가의 보완을 위해서는 선호도 선상에서 적당한 거리로 떨어져 있는 곡들을 선정하여 해당 곡들에 대한 순위를 매기도록 요청하는 것이 적합할 것으로 보인다.

위에서 서술한 평가 설계의 보완 방향과 별개로, 암시적 피드백 특성의 가공 과정 자체로 인해 첫 번째 평가 문항과 같은 방식으로 특성을 선호도 모델로 사용하여 곡들을 정렬한 결과가 뚜렷한 특징을 보이지 않을 가능성도 존재한다. 재생횟수-평균 가중 재생을 곱 특성은 시간에 따라 감소할 수 있는 각 아이템에 대한 사용자의 선호도를 반영할 수 있도록 고안되었으며, 이에 따라 어떠한 곡에 대한 과거와 현재의 사용자의 선호도 차이가 크다고 가정했을 때 특성값은 과거와 현재의 선호도의 중간 정도를 나타내게 된다. 이러한 특징을 가진 값을 기준으로 곡들을 정렬한 결과만을 보았을 때 사용자 입장에서 직관적이지 않을 수 있으며, 이 때문에 재생횟수-평균 가중 재생을 곱 특성의 선호도 모델로써의 성능 검증을 위한 첫 번째 문항에서의 평가 결과가 낮은 수치로 나타났을 가능성이 높다.

그럼에도 불구하고 사용자 평가 결과에 의하면 재생횟수-평균 가중 재생을 곱 특성은 협업적 음악 추천 알고리즘에 적용되었을 때 가장 높은 전체 만족도 점수를 받은 곡 추천 결과를 도출했으며, 가장 새로운 곡을 추천하였다. 본 연구에서 추천 결과의 평가를 위해 사용된 관련성, 새로움, 우연성을

포함한 여러 평가 기준은 추천 시스템과 추천 대상 사용자의 목적에 따라 해석이 달라질 수 있다. 그러나 많은 선행 연구에서는 추천 시스템이 사용자가 이미 익숙하거나 인지도/인기도가 높은 아이টে임을 추천하는 것은 바람직하지 않음을 지적하였으며, 이 때문에 새롭거나(novel) 롱테일(long-tail) 선상에 있는 아이টে임을 추천하는 현상에 대해 긍정적으로 해석해왔다(Chou et al., 2015; Jannach et al., 2013). 이러한 동향을 고려했을 때, 다른 리스트에 비해 더 새로우면서도 만족도 점수는 높은 곡 추천 리스트를 도출한 재생횟수-평균 가중 재생율 곱 특성은 추천 과정에 긍정적인 영향을 미쳤다고 할 수 있다.

결론적으로는 사용자 평가 설계의 여러 한계점으로 인해 참여자들의 곡 선호도 선상에서 극상위를 제외한 중간 및 하위 순위의 곡들에 대한 암시적 피드백 특성들의 선호도 모델링 성능을 비롯한 전반적인 성능을 구체적으로 관찰하지 못하였다. 그러나 평가 결과에 의하면 참여자들은 음악 선호도 모델 평가에서는 누적 가중 재생율 특성을, 음악 추천 결과 평가에서는 재생횟수-평균 가중 재생율 곱 특성을 높게 평가하였다. 이러한 결과를 통해 사용자의 선호도가 높은 극상위 순위의 곡들에 대한 선호도만을 잘 반영하는 암시적 피드백 모델보다는, 가중 재생율 개념을 도입하여 중간 순위의 곡을 비롯하여 사용자가 어떤 곡을 불호했는지까지도 잘 추정할 수 있는 모델이 좋은 음악 추천으로 이어질 가능성이 높다는 점을 시사할 수 있다.

5.3.2. 음악 콘텐츠 소비 경향 차이에 따른 선호도 모델링 방식

재생횟수-평균 가중 재생율 곱 특성이 추천 과정에 긍정적인 영향을 미친 이유 중 하나는 아마도 곡에 대한 선호도를 모델링하는 과정이 음악 도메인 콘텐츠의 소비 경향을 더 잘 반영했기 때문일 것으로 추측할 수 있다. 초기의 협업적 필터링 추천 방식에 대한 연구는 이메일, 혹은 뉴스 콘텐츠에 초점을 맞췄으며(Goldberg et al., 1992; Konstan et al., 1997), 추천 분야의 연구를 위해 최근 가장 많이 설명되고 사용되는 데이터 도메인 중

하나는 영화이다. 음악 도메인의 콘텐츠와 타 도메인의 콘텐츠와의 가장 큰 차이점 중 하나는 소비 시간이 비교적 짧다는 특징과, 상품이면서도 예술 작품이라는 특징이 결합함으로써 나타나는 반복적인 소비 경향이다. 사용자들은 영화에 비해 훨씬 많은 수의 음악 콘텐츠를 접하고 소비하게 되며, 반복적으로 소비한 음악 콘텐츠에 대해서는 다양한 방식으로 선호도의 변화가 자주 일어날 가능성이 높아진다.

기존 재생횟수 특성은 이와 같은 선호도를 누적의 형태로 모델링하며, 이러한 방식으로 인해 소비한 지 오래된 아이템일수록 높은 선호도를 가질 수밖에 없는 모델이 된다. 본 연구에서 제안한 특성 중 누적 가중 재생을 역시 재생횟수와 유사하게 소비를 누적하는 대신 매 소비를 더 높은 해상도로 모델링하며, 사용자 평가 결과를 통해 이러한 방식이 곡들의 선호도를 구별하고 정렬하는 데에 더 효과적임을 알 수 있다.

반면 재생횟수-평균 가중 재생을 곱 특성의 경우 누적 방식이 아닌, 특정 곡에 대한 과거의 선호도를 고려하면서도 청취 패턴 혹은 선호도의 변화에 따라 부정적인 소비 경험 역시 반영할 수 있다는 특징을 지니며, 이는 앞서 설명한 음악 도메인 콘텐츠에 대한 선호도를 모델링하는 데에 더 적절한 방식이라고 할 수 있다. 이와 같은 가정은 음악 추천 알고리즘에서의 성능 평가 단계를 통해 어느 정도 검증되었으며, 참여자들의 피드백을 통해서도 음악 도메인 콘텐츠에 대해서는 단순한 누적 방식보다는 개인화된 기준에 따라 각 곡에 대한 선호도 변화를 반영할 수 있는 모델이 필요함을 확인할 수 있었다. 다음은 참여자들에게 재생횟수를 기반으로 정렬하는 방식을 택하고 있는 Last.fm 의 “Top tracks” 차트가 자신의 전반적인 선호도 순위를 얼마나 잘 표현하는지를 묻는 사전 설문 의 한 질문에 대한 일부 답변을 나타낸다.

“제가 (Last.fm을 통해) 처음 scrobbling 하기 시작했을 때는 곡 수가 별로 많지 않았기 때문에, 저의 초기 청취 습관들이 불가

피하게 지나치게 반영된 것 같습니다.”

“몇몇은 여전히 제가 선호하는(favorite) 곡들이지만, 아직도 그 곡들이 최고(top)이라고 할 수는 없을 것 같아요.”

“제가 한 때 집착했던 곡/앨범을 보여주는 하지만, 제가 실제 Top tracks 라고 여기는 곡 리스트와 유사하지는 않아요.”

“차트에는 제가 이제는 많이 듣지 않는 곡들이 있어요.”

위와 같은 해석으로 미루어 보았을 때, 사용자들이 청취하는 음악 라이브러리의 크기가 작고, 자신이 잘 인지하고 있는 작은 규모의 곡들을 반복적으로 청취하는 성향일수록 재생횟수 혹은 본 연구를 통해 제안한 특성 중 누적 가중 재생율과 같이 소비를 누적하는 방식으로 모델링된 암시적 피드백 특성이 사용자의 선호도를 나타내기에 충분할 것임을 예측할 수 있다. 반대로, 청취하는 음악 라이브러리의 크기가 빠르게 성장하고, 이 때문에 곡 레벨에서의 선호도가 시기에 따라 서로 다르게 변화할 가능성이 높은 사용자들의 경우에는 본 연구에서 제안한 특성 중 재생횟수-평균 가중 재생율 곱과 같이 개인화된 기준을 기반으로 곡 레벨에서의 선호도 감소 역시 반영할 수 있는 암시적 피드백 모델을 적용하는 것이 적합할 것이다.

제 6 장 결 론

제 1 절 결론 및 연구 의의

본 연구에서는 기존 음악 추천 시스템에서 암시적 피드백으로 대부분 채택되고 있는 재생횟수의 한계를 보완하고, 실제 사용자의 곡 선호도를 보다 잘 반영할 수 있는 암시적 피드백 특성을 발굴하기 위한 목적으로 음악 청취 이벤트의 재생율을 활용하여 누적 가중 재생율과 재생횟수-평균 가중 재생율이라는 두 가지 암시적 피드백 특성을 제안하였다. 이 후 제안된 암시적 피드백 특성이 기존의 특성에 비해 사용자들의 곡 선호도를 더 잘 반영할 수 있는지와, 실제 음악 추천 알고리즘에 적용되었을 때의 추천 결과에 차이가 있는지를 사용자 평가를 통해 검증하였다. 약 14명의 참여자들의 평가 데이터를 분석한 결과, 통계적으로 유의하지는 않았으나 제안한 특성 중 누적 가중 재생율을 사용하여 사용자들이 청취한 곡을 정렬했을 때 사용자들의 실제 선호도 순위와 가장 유사하다는 평가를 받으며, 재생횟수-평균 가중 재생율의 경우 사용자의 실제 선호도 순위와 비교해서 가장 낮은 유사도 점수를 받았다. 음악 추천 알고리즘에서의 경우, 제안한 특성 중 재생횟수-평균 가중 재생율 특성을 암시적 피드백으로 사용했을 때 도출된 추천 리스트에 대한 사용자들의 평점이 가장 높았으며, 추후 실험 보완을 통해 통계적으로도 유의한 차이를 이끌어낼 수 있는 가능성을 보였다. 또한, 재생횟수-평균 가중 재생율 특성을 사용했을 때 유의한 차이로 가장 관련성(relevance)이 낮으면서 거의 유의한 차이로 가장 새로운(novel) 곡들을 추천하였고, 우연성(serendipity) 척도에 대한 점수 역시 재생횟수에 비해 높은 축에 속했다. 누적 가중 재생율 특성의 경우, 유의한 차이로 다른 특성에 비해 가장 관련성(relevance)이 높은 곡들을 추천하였다. 이러한 결과를 통해 재생율을 암시적 피드백 특성에 도입하는 것이 재생횟수에 비해 음악 추천 결과에 대한 사용자들의 높은 만족도로 이어질 가능성이 있으며, 특히 누적 가중 재생율의 경우 관련성이 높은 곡 추천 결과를, 재생횟수-평균 가중 재생율의 경우 새로운 곡 추천 결과를 도출할 수 있음을 관찰하였다.

위와 같이 상반되는 음악 선호도 모델에 대한 평가와 음악 추천 알고리즘에서의 성능 평가 결과는 암시적 피드백이 사용자들의 곡 선호도를 잘 반영할수록 음악 추천 알고리즘에서의 성능 역시 높아질 것이라는 연구의 기본적인 가정을 반하는 것처럼 보였으나, 평가 설계의 한계를 기반으로 고찰해보았을 때 본 사용자 평가를 통해서 제대로 평가하지 못한 선호도 선상의 극상위를 제외한 중간 및 하위 순위의 아이템들에 대한 암시적 피드백 특성들의 선호도 모델링 성능이 추천 성능과 높은 상관관계가 있을 가능성이 있음을 알 수 있었다. 또한 사용자 평가를 통해서 관찰하지 못했으나, 제안한 피드백 특성의 가공 방식, 재생횟수 기반 곡 정렬 방식에 대한 평가 참여자들의 피드백과 음악 도메인 콘텐츠의 소비 경향의 특징을 견주어 고찰해보았을 때 재생횟수-평균 가중 재생을 곱 특성과 같이 개인화된 기준을 기반으로 곡 레벨에서의 선호도 감소 역시 반영하는 암시적 피드백 모델이 음악 도메인 콘텐츠에 대한 선호도 모델링에는 더 적합할 것이라고 추론할 수 있었다.

본 연구는 음악 추천 시스템 연구 분야에서 최근 증가하고 있는 사용자-중심적(user-centric)인 연구 동향에 맞추어, 사용자들의 음악 청취 데이터를 기반으로 기존의 암시적 피드백 특성으로 오랜 기간 사용되어 온 재생횟수보다 음악 선호도를 보다 잘 반영할 수 있는 새로운 특성들을 발굴하려는 시도를 했다는 점에서 의의가 있다. 또한 본 연구는 암시적 피드백 모델이 사용자들의 실제 선호도를 잘 반영할 수 있는지 검증하기 위해 새로운 평가 프레임워크를 설계하여 적용했다는 점에서, 추후 유사한 연구 문제의 평가 프레임워크에 참고가 될 수 있을 것이다. 마지막으로, 본 연구에서는 음악 도메인 콘텐츠의 소비 경향을 적절히 반영하는 선호도 모델을 구현하는 것이 추천 성능에 긍정적인 영향을 미칠 수 있는 가능성을 보이고, 이에 따라 도메인 특성을 고려한 선호도 모델링 연구의 중요성을 시사했다는 점에서 연구 의의가 있다.

제 2 절 연구의 한계 및 향후 연구

본 연구에서는 암시적 피드백의 곡 선호도 모델로서의 성능을 평가하기 위해 이전에 시도되지 않았던 평가 프레임워크를 적용했기 때문에 이에 따른 한계점이 존재했다. 먼저 5장의 고찰을 통해 설명한 바와 같이, 평가를 위해 참여자들에게 주어진 곡들이 선호도 선상에서 가까이 위치하기 때문에 이들 간의 선호도 우열을 평가하기가 모호하다는 한계점이 있었다.

이 외에도 사용자 평가 방식을 택했기 때문에 제안한 암시적 피드백 특성의 튜닝(tuning) 과정을 거치지 못했으며, 이는 제안한 암시적 피드백 특성이 추가적으로 최적화될 수 있을 여지가 있음을 의미한다. 사용자 평가에 참여한 사용자들의 수 역시 유의미한 결론 도출을 위해 부족했을 가능성 역시 존재한다. 또한 다른 연구에서 겪는 문제와 마찬가지로, 사용자 평가를 위해 추천 알고리즘에 사용된 데이터와 평가 참여자들이 평가를 수행한 시점이 일치하지 않기 때문에 데이터셋이 수집된 시기 이후 발매된 곡들은 추천될 수 없었다는 한계점이 있다. 마지막으로, 기존의 데이터셋에 존재하지 않았던 곡 지속 시간 데이터를 수집하는 과정에서 누락된 곡들이 많았기 때문에, 이러한 곡들에 대한 사용자들의 청취 데이터를 분석하거나 해당 곡들을 추천하는 것이 불가능했다는 점 역시 본 연구의 큰 한계점 중 하나이다.

위와 같은 한계점에 따라, 추후 연구에서는 누락 데이터를 보완하고, 선호도 모델 평가를 위한 프레임워크 중 평가 대상 곡 선정 과정을 개선하여 재생율을 활용한 추가적인 암시적 피드백 특성을 발굴하는 연구를 시도할 수 있을 것이다. 이와 더불어, 본 연구의 결과에서는 관찰할 수 없었던 암시적 피드백 모델과 음악 추천 알고리즘의 성능과의 상관관계를 분석함으로써, 곡에 대한 선호도를 암시적 피드백으로 모델링하는 방식이 추천 결과의 다양한 척도에 어떠한 방식으로 영향을 미치는지 연구해볼 수 있을 것으로 예상된다.

참고 문헌

- Adomavicius, G., & Tuzhilin, A. (2005). 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*, 17(6), 734–749.
- Aggarwal, C. C. (2016). *Recommender systems* (pp. 1–28). Springer International Publishing.
- Peynirciđlu, Z. F., Tekcan, A. Í., Wagner, J. L., Baxter, T. L., & Shaffer, S. D. (1998). Name or hum that tune: Feeling of knowing for music. *Memory & cognition*, 26(6), 1131–1137.
- Amatriain, X., Pujol, J. M., & Oliver, N. (2009, June). I like it... i like it not: Evaluating user ratings noise in recommender systems. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 247–258). Springer, Berlin, Heidelberg.
- Amatriain, X., Pujol, J. M., Tintarev, N., & Oliver, N. (2009, October). Rate it again: increasing recommendation accuracy by user re-rating. In *Proceedings of the third ACM conference on Recommender systems* (pp. 173–180). ACM.
- Andric, A., & Haus, G. (2006). Automatic playlist generation based on tracking user's listening habits. *Multimedia Tools and Applications*, 29(2), 127–151.
- Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011, October). The Million Song Dataset. In *Ismir* (Vol. 2, No. 9, p. 10).
- Cao, L., & Guo, M. (2008, December). Consistent music recommendation in heterogeneous pervasive environment. In *Parallel and Distributed Processing with Applications, 2008. ISPA'08. International Symposium on* (pp. 495–501). IEEE.

- Celma Herrada, Ò. (2009). Music recommendation and discovery in the long tail.
- Celma, O. (2010). Music recommendation. In *Music recommendation and discovery* (pp. 43–85). Springer, Berlin, Heidelberg.
- Chou, S. Y., Yang, Y. H., & Lin, Y. C. (2015, June). Evaluating music recommendation in a real-world setting: On data splitting and evaluation metrics. In *Multimedia and Expo (ICME), 2015 IEEE International Conference on* (pp. 1–6). IEEE.
- David, N. (1997, November). Implicit Rating and Filtering. In *5th DELOS Workshop on Filtering and Collaborative Filtering (ERCIM)*, Budapest, Hungary (pp. 31–36).
- Deshpande, M., & Karypis, G. (2004). Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1), 143–177.
- Farrahi, K., Schedl, M., Vall, A., Hauger, D., & Tkalcic, M. (2014, October). Impact of Listening Behavior on Music Recommendation. In *ISMIR* (pp. 483–488).
- Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., & Reiterer, S. (2013). Toward the next generation of recommender systems: applications and research challenges. In *Multimedia services in intelligent environments* (pp. 81–98). Springer, Heidelberg.
- Gadanho, S. C., & Lhuillier, N. (2007, October). Addressing uncertainty in implicit preferences. In *Proceedings of the 2007 ACM conference on Recommender systems* (pp. 97–104). ACM.

- Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61–70.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157–1182.
- Hanna, P. (2017). Considering Durations and Replays to Improve Music Recommender Systems. arXiv preprint arXiv:1711.05237.
- Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999, August). An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 230–237). ACM.
- Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, 22(1), 5–53.
- Hu, Y., Koren, Y., & Volinsky, C. (2008, December). Collaborative filtering for implicit feedback datasets. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on* (pp. 263–272). Ieee.
- Jannach, D., Lerche, L., Gedikli, F., & Bonnin, G. (2013, June). What recommenders recommend—an analysis of accuracy, popularity, and sales diversity effects. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 25–37). Springer, Berlin, Heidelberg.
- Jawaheer, G., Szomszor, M., & Kostkova, P. (2010, September). Characterisation of explicit feedback in an online music recommendation service. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 317–320). ACM.

- Jawaheer, G., Szomszor, M., & Kostkova, P. (2010, September). Comparison of implicit and explicit feedback from an online music recommendation service. In proceedings of the 1st international workshop on information heterogeneity and fusion in recommender systems (pp. 47–51). ACM.
- Jawaheer, G., Weller, P., & Kostkova, P. (2014). Modeling user preferences in recommender systems: A classification framework for explicit and implicit user feedback. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 4(2), 8.
- Johnson, C. C. (2014). Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, 27.
- Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(1), 2.
- Kelly, D., & Teevan, J. (2003, September). Implicit feedback for inferring user preference: a bibliography. In *Acm Sigir Forum* (Vol. 37, No. 2, pp. 18–28). ACM.
- Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3), 77–87.
- Kordumova, S., Kostadinovska, I., Barbieri, M., Pronk, V., & Korst, J. (2010, June). Personalized implicit learning in a music recommender system. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 351–362). Springer, Berlin, Heidelberg.
- Lee, T. Q., Park, Y., & Park, Y. T. (2008). A time-based approach to effective recommender systems using implicit feedback. *Expert systems with applications*, 34(4), 3055–3062.

- Liu, Y., Zhao, P., Sun, A., & Miao, C. (2015, July). A Boosting Algorithm for Item Recommendation with Implicit Feedback. In *IJCAI* (Vol. 15, pp. 1792–1798).
- Najafabadi, M. K., & Mahrin, M. N. R. (2016). A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback. *Artificial intelligence review*, 45(2), 167–201.
- Parra, D., Karatzoglou, A., Amatriain, X., & Yavuz, I. (2011). Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping. *Proceedings of the CARS-2011*.
- Parra-Santander, D., & Amatriain, X. (2011). Walk the Talk: Analyzing the relation between implicit and explicit feedback for preference elicitation.
- Schedl, M., Gómez, E., & Urbano, J. (2014). Music information retrieval: Recent developments and applications. *Foundations and Trends® in Information Retrieval*, 8(2–3), 127–261.
- Schedl, M. (2016, June). The lfm-1b dataset for music retrieval and recommendation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval* (pp. 103–110). ACM.
- Song, Y., Yang, P., Zhang, C., & Ji, Y. (2015, August). Implicit feedback mining for recommendation. In *International Conference on Big Data Computing and Communications* (pp. 373–385). Springer, Cham.
- Su, J. H., Chang, W. Y., & Tseng, V. S. (2013). Personalized music recommendation by mining social media tags. *Procedia Computer Science*, 22, 303–312.
- Yoshii, K., Goto, M., Komatani, K., Ogata, T., & Okuno, H. G. (2006, October). Hybrid Collaborative and Content-based Music

Recommendation Using Probabilistic Model with Latent User Preferences.
In ISMIR (Vol. 6, p. 7th).