

# 안드로이드 환경에서 협업필터링을 이용한 공연 추천 시스템

\*장태훈, \*신해란, \*임성훈, \*박두순  
 \*순천향대학교 컴퓨터소프트웨어공학과  
 e-mail : wkdxogns23@naver.com

## Performance Recommendation System using Collaborative Filtering in Android

Tae-Hoon Jang, HaeRan Shin, Doo-Soon Park, Sung-Hun Lim  
 Dept. of Computer Software Engineering, Soonchunghyang University

### 요 약

현대사회가 발전함으로 인해 사람들의 질이 높아졌고 보다 행복한 삶을 찾는 사람들이 많아졌다. 그로인해 공연 문화의 시장은 커졌고 공연의 종류와 수가 매우 많아졌다. 이에 따라 사용자에게 만족되는 조건과 가격의 공연을 추천하기란 어려운 일이다. 따라서 본 논문에서는 개인화 요인과 협업 필터링 방법을 이용하여 사용자에게 보다 적합한 공연을 추천하며 협업필터링의 큰 단점 중 하나인 희소성 문제에 대해 보다 나은 추천 시스템을 제안한다.

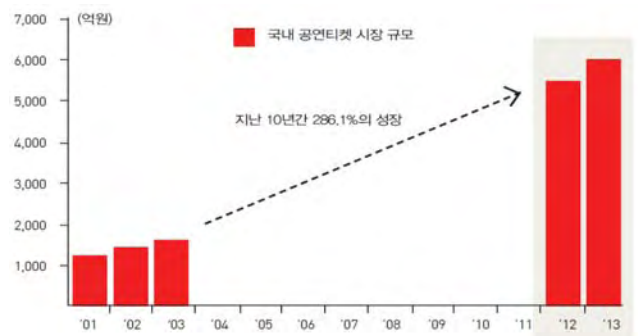
### 1. 서론

최근 현대사회는 빠른 속도로 발전하고 있다. 이로 인해 사람들은 점차 더 나은 질의 생활과 행복한 삶에 대한 욕망이 점차 커지고 있다. 이러한 상황에 발 맞춰 여가 문화는 다양해지고 있는 추세이다. 또한 많은 여가 활동 중 공연 문화를 즐기는 인구는 꾸준히 증가하고 있다. 이에 따라 공연 시장의 크기는 매우 커지고 있으며 과거 공연의 종류가 연극, 희극, 뮤지컬, 연주회 등에 국한되어 있었다면 지금의 공연의 종류는 콘서트, 개그, 관객모독, 난타 등 매우 다양하고 그 수 또한 매우 많아지고 있다. 하지만 이러한 공연의 특성상 특정 지역에서만 행해지며 가격과 스타일 또한 가지각색이기 때문에 사용자를 만족 시키는 공연을 추천하는 것은 쉬운 일이 아니다.

2014년 기준 국내 공연시장 규모 표는 (그림 1)과 같고 2001년 ~ 2013년까지의 국내 공연티켓 시장의 증가율은 (그림 2)와 같다.

2014년 기준		
공연시설 수	↑	1,034개
공연장 수	↑	1,280개
공연시설 종사자 수	↑	1만 2,669명
공연단체 수	↑	2,284개
공연단체 종사자 수	↑	5만 5,858명
공연 건수	↑	4만 7,489건
공연 횟수	↑	20만 228회

(그림 1) 2014년 기준 국내 공연시장 규모[1]



(그림 2) 2006년 ~ 2014년 국내 공연티켓 시장 규모[2]

현재 공연을 알아보는데 도움을 주는 옥션티켓(ticket.auction.co.kr), 네이버 공연검색(https://search.naver.com)과 같은 홈페이지가 존재하지만 단지 자신이 원하는 공연을 직접 찾은 후 그 공연의 설명을 보거나 티켓을 예매하는

※ 본 연구는 NRF-2017R1A2B1008421에 의해 지원되었음.

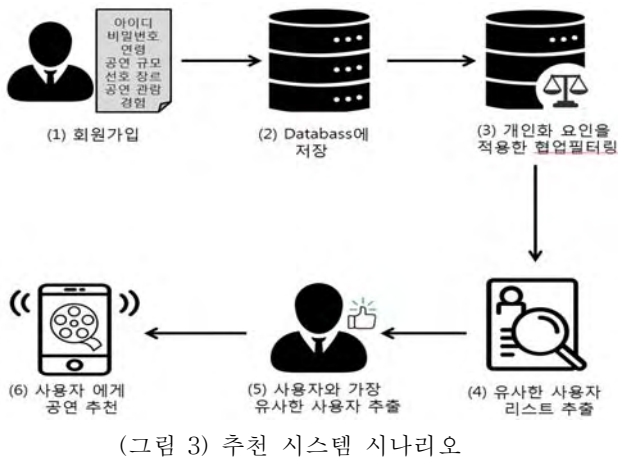
것이 기능의 전부다. 때문에 사용자들은 일일이 공연들을 검색한 후 후기를 읽어 가며 자신의 취향에 맞는 공연을 찾아야하는 불편함이 있다. 또한 종류가 매우 많기 때문에 사용자는 공연의 정보를 검증하는데 많은 시간을 소비하고 있다.

본 논문에서는 기존의 사용자 기반 협업필터링의 문제 중 희소성 문제에 대해서 보다 나은 방안을 제시한다. 협업 필터링에 대한 데이터가 적을 시 발생하는 희소성 문제를 개인화 요인을 바탕으로 협업 필터링을 사용하여 사용자들 사이에 유사성을 판별한다. 이렇게 형성된 데이터를 기반으로 사용자의 취향에 가장 적합한 공연을 추천한다. 또한 안드로이드 환경에서 구현하여 편의성을 높인 공연 추천 시스템을 제안한다.

## 2. 공연 추천 시스템의 구성

협업 필터링의 주요 기술로는 유사한 성향을 가진 사람들을 구분하고, 해당 성향의 사람들이 좋아하는 것을 추천하는 방식인 사용자기반 협업 필터링과, 사람이 아닌 아이템을 비교하여 유사한 아이템을 추천하는 아이템기반 협업 필터링 기법이 있다. 본 논문에서는 사용자기반 협업 필터링을 사용하여 유사도를 측정하였고 이를 통해서 근접이웃을 구성하였다. 하지만 큰 단점이 있다면 기존의 사용자가 아닌 새로운 사용자에 대한 추천에 대해서는 제대로 된 추천이 어렵다는 것이다. 그러므로 본 논문에서는 이러한 문제 중 하나인 희소성 문제에 있어 보다 적절한 추천을 할 수 있는 방법을 제시한다.

본 논문에서 구현하게 될 공연 추천 시스템의 시나리오는 (그림 3)과 같다.



(그림 3)의 추천 시스템 시나리오의 설명은 다음과 같다.  
 (1) 먼저 사용자는 추천 시스템을 사용하기에 앞서 안드로이드 어플리케이션을 통해 회원가입을 필수로 해야 하고 개인화 요인을 입력하게 된다. 이 때 사용자는 연령, 공연 규모, 공연 관람 경험, 선호 장르를 입력하게 되는데 이러한 데이터는 (2) 서버에 위치한 데이터베이스에 저장한다.

(3) 그 후 협업 필터링의 문제 중 하나인 희소성 문제를 해결하기 위해 데이터베이스에 저장된 사용자의 개인화 요소에 가중치를 부여하고 기존의 이용 중인 사용자와의 유사도를 피어슨 상관관계 공식을 이용하여 협업필터링을 통해 추출한다.

협업 필터링이란, 사용자들의 선호도와 관심 표현을 바탕으로 선호도, 관심도가 비슷한 사용자들을 식별해 내는 방법으로 과거에 이용한 콘텐츠가 비슷하다면 사용자 간에 유사한 성향을 가지고 있다고 판단하고 그 근거를 토대로 이루어진다.[3]

(4) 협업필터링을 통해 추출된 데이터 중 사용자와 유사한 사용자 N명을 근접이웃으로 구성하고 (5) 유사도가 가장 높은 사용자 5명을 추출해 TOP\_5 List를 구성한다. (6) 이때 만들어진 TOP\_5 List의 사용자가 추천하는 공연을 사용자에게 안드로이드 어플리케이션을 통해서 추천 한다.

개인화 요인으로 사용하게 될 요소들은 아래의 (표 1)과 같다.

(표 1) 개인화 요인 사용 요인

번호	연령	공연 규모	공연 관람 경험	선호 장르
1	10대	10명 이하	무	연극
2	20대	50명 이하	1회	희극
3	30대	100명 이하	5회 이하	뮤지컬
4	40대	200명 이하	10회 이하	연주회
5	50대	500명 이하	10회 이상	난타
6	60대	500명 이하		개그
7	70대 이상			콘서트
8				관객모독
9				난타
10				마임
11				마술

위의 (표 1)과 같이 분류한 요소는 두 가지 논문에서 개인화 요소로 사용했던 개인화 요소인 '연령대, 공연규모, 소득분위, 선호지역, 성별, 국내/국외 공연, 선호 장르, 관람 경험 유/무' 중에서[4][5] '연령, 공연 규모, 공연 관람 경험, 선호 장르'를 사용하여 추천하였을 때 높은 수치의 유사한 사용자들이 다량으로 등장했고 가장 적절한 추천이 가능했기에 선정 된 개인화 요소다. (표 2)는 같은 데이터에서 개인화 요인의 종류와 수를 바꾸어 추천했을 때의 유사도를 높은 순으로 정렬한 상위 10개를 보여준다.

(표 2) 개인화 요인에 따른 추천 결과 유사도 상위 10개의 데이터

번호	사용한 개인화 요인	유사도
1	연령, 공연규모, 공연 관람 경험, 선호 장르	0.879
2	연령, 공연규모, 소득분위, 선호지역	0.856
3	연령, 공연규모, 공연 관람 경험, 선호 장르	0.801
4	소득분위, 선호지역	0.724
5	국내/국외 공연, 관람 경험 여부, 소득 분위	0.716
6	연령, 공연규모, 공연 관람 경험, 선호 장르	0.712
7	연령, 공연규모, 공연 관람 경험, 선호 장르	0.701
8	연령, 선호 장르, 성별	0.613
9	소득분위, 성별	0.608
10	공연 규모, 선호 장르, 성별	0.587

표를 보면 알 수 있듯 유사도가 높게 나온 상위 10개의

데이터 중 4개의 항목에서 ‘연령, 공연 규모, 공연 관람 경험, 선호 장르’를 사용하여 추천한 결과라는 것을 볼 수 있었고 이러한 4가지 항목을 사용했을 때 보다 유사한 사용자를 찾을 수 있는 확률이 늘어났다. 이러한 이유에서 위의 4가지 개인화 요인을 기반으로 협업필터링을 통해 공연을 추천해주는 절차를 갖는다.

### 3. 소설 추천 시스템의 구현

본 논문에서 구현한 공연 추천 시스템은 협업 필터링의 희소성 문제를 해결하기 위해 회원가입이 필수로 되어야 이용이 가능하다. 회원가입을 통해 사용자는 개인화 요인을 입력하게 되는데 요소는 아이디, 비밀번호, 연령, 공연 규모, 공연 관람 경험, 선호 장르로 이루어져있다. 안드로이드 스튜디오에서 에뮬레이터를 이용하여 실행시킨 회원가입의 양식은 (그림 4)와 같다.



(그림 4) 회원 가입 양식

다음과 같은 양식에 따라서 사용자가 입력한 정보는 회원 관리 테이블에 저장되고, 회원관리 테이블은 아이디, 연령, 공연 규모, 공연 관람 경험, 선호 장르 순으로 저장된다. 회원관리 데이터베이스는 (그림 5)와 같다.

ID	BIRTH	SCALE	EXPERIENC	GENRE
20134036	2	2	1	4
20154203	1	4	2	3
20110542	4	1	3	1
20160512	1	3	4	3
20120420	2	4	1	7
20170212	1	6	5	3
20150433	2	2	1	11
20154006	3	5	3	4

(그림 5) 회원정보 데이터베이스

본 논문에서는 사용자가 회원가입 할 때 입력한 개인화 요인에 연령(40%), 공연 규모(25%) 공연 관람 경험(25%), 선호 장르(10%)로 가중치를 부여하였다. (표 3)은 가중치를 바꿔가며 추천했을 때의 유사도가 높은 순으로 정렬하였을 때의 상위 10개의 데이터다.

(표 3) 가중치에 따른 추천 결과 유사도 상위 10개의 데이터

번호	사용한 개인화 요인	유사도
1	연령(40), 공연규모(25), 공연 관람 경험(25), 선호 장르(10)	0.911
2	연령(40), 공연규모(25), 공연 관람 경험(25), 선호 장르(10)	0.884
3	연령(30), 공연규모(30), 공연 관람 경험(30), 선호 장르(10)	0.862
4	연령(25), 공연규모(25), 공연 관람 경험(25), 선호 장르(25)	0.815
5	연령(50), 공연규모(15), 공연 관람 경험(15), 선호 장르(20)	0.798
6	연령(40), 공연규모(20), 공연 관람 경험(20), 선호 장르(20)	0.755
7	연령(25), 공연규모(25), 공연 관람 경험(25), 선호 장르(25)	7.08
8	연령(35), 공연규모(35), 공연 관람 경험(20), 선호 장르(10)	6.987
9	연령(30), 공연규모(30), 공연 관람 경험(20), 선호 장르(20)	6.856
10	연령(40), 공연규모(25), 공연 관람 경험(25), 선호 장르(10)	6.422

(표 3)을 보면 알 수 있듯 가중치를 바꿔가며 추천해 본 결과 본 논문에서 제시하는 가중치를 주었을 때 유사도가 가장 높은 결과들이 많이 나왔고 다음과 같은 결과로 선정된 4가지 요소의 가중치를 결정했다.

이러한 데이터를 토대로 사용자와 다른 사용자간의 유사도를 피어슨 상관계수를 통해 구한다. 피어슨 상관계수는 -1부터 1의 값을 가지며, 관계가 크면 1에 가까워지고 서로의 유사함이 적으면 -1에 가까워진다. 피어슨 상관계수를 이용한 수식은 아래 (그림 6)과 같다.[6]

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

(그림 6) 피어슨 상관계수 공식

이러한 피어슨 상관계수를 이용하여 사용자의 유사도 측정의 한 결과를 토대로 근접이웃을 구성하고 사용자에게 대한 Top-N 리스트의 일부를 나타내면 (그림 7)과 같다.

ID	compare user	similarity
20134036	20134036	1
	20154203	0.294
	20110542	0.134
	20160512	0.327
	20120420	0.674
	20170212	0.257
	20150433	0.837
	20154006	0.424

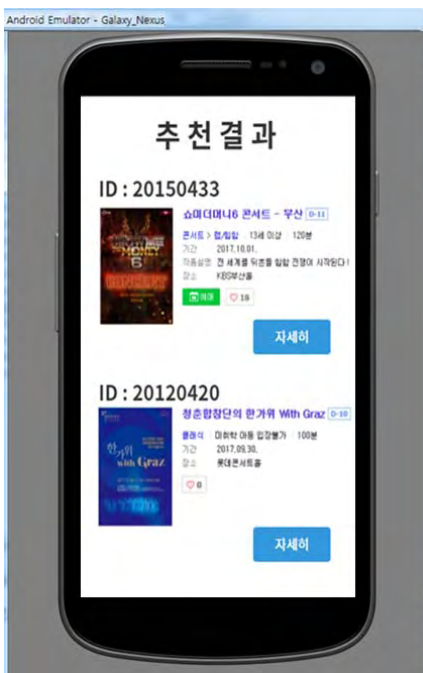
(그림 7) 사용자에게 대한 Top-N List

Top-N List의 근접이웃 중 유사도가 가장 높게 측정된 5명의 사용자를 다시 추출하게 되는데 이 때 추출된 5명의 사용자가 추천한 공연을 사용자에게 추천하게 된다. (그림 8)은 이렇게 추천된 5명의 사용자가 추천한 공연이다.

ID	similarity	result
20150433	0.837	청춘합창단 제3회 정기 연주회
20120420	0.674	쇼미더머니6 콘서트
20134245	0.664	행오버
20152204	0.642	김종욱 찾기
20105542	0.637	국립발레단 안나 카레니나
20145732	0.597	두 여자

(그림 8) 사용자와 가장 유사한 5명의 공연 추천 리스트

안드로이드 구현 환경에는 Activity와 View의 두 부분으로 분류가 되는데, Activity에서 Java API 프레임워크를 이용하여 UI/UX를 동작시키도록 구현하였고, View에서는 UI/UX를 동작시키기 위한 디자인 및 버튼 등을 구현하였다. 때문에 추천 결과는 Activity와 View를 통해 구현한 안드로이드 어플리케이션을 통해 확인 할 수 있다. (그림 11)는 안드로이드 스튜디오에서 에뮬레이터를 통해 실행시킨 어플리케이션 결과 화면을 캡처한 것이다.



(그림 11) 추천 결과 화면

요인의 가중치를 부여하고 이를 피어슨 상관계수로 계산하여 보다 정확한 결과를 추출하는 시스템이다. 또한 사용자가 보다 편리하게 추천결과를 볼 수 있도록 안드로이드 환경에서 구현하였다.

향후 연구 과제는 보다 정확한 추천을 위해 더 많은 종류의 개인화 요소를 검증 해보고 더욱 적합한 개인화 요소와 가중치를 찾는다면 더욱 발전된 시스템이 될 것이다.

### 참고문헌

- [1] “공연시장분석”-문화체육관광부 2015.02.35
- [2] “KOCCA 통계로 보는 콘텐츠산업”-한국콘텐츠진흥원 <http://koreancontent.kr> 2015.01.06.
- [3] 김영아, 박두순, “협업 필터링 기반 드라마 추천 시스템”, 순천향대학교 컴퓨터소프트웨어공학과 2013.11
- [4] 정연우, 박두순, “개인화 요인과 협업필터링 기반의 도서 추천 시스템”, 순천향대학교 컴퓨터소프트웨어공학과 2015.09
- [5] 장태훈, 김한이, 박두순, “클라우드 컴퓨팅 환경에서 협업 필터링과 개인성향을 이용한 개인화 소셜 추천 시스템”, 순천향대학교 컴퓨터소프트웨어공학과 2016.03
- [6] 장태훈, Khamphaphone, 박두순, “ 협업 필터링과 개인 성향을 이용한 개인화 자취방 추천 시스템”, 순천향대학교 컴퓨터소프트웨어공학과 2016.09

## 4. 결론

본 논문에서는 수많은 공연 중에서 사용자의 기호에 가장 적합한 공연을 보다 정확하게 추천해주는 시스템을 구현했고 협업필터링의 큰 문제점인 희소성 문제를 개인화