

# 개인화 요인과 협업필터링을 이용한 드라마 추천 시스템

김민기\*, 박두순\*

\*순천향대학교 컴퓨터소프트웨어공학과

e-mail : kmk6715k@naver.com

## Drama Recommendation System Using Personal Elemnets and Collaborative Filtering

\*Min-Ki Kim, \*Doo-Soon Park

\*Dept. of Computer Software Engineering, SoonChunHyang University

### 요 약

최근 한국의 드라마가 국내뿐만 아니라 국외에서도 인기를 끌고 있다. 이로 인해 지상파 채널을 비롯해 중편 채널과 많은 케이블채널에서 전보다 많은 드라마가 등장하고 있으며 드라마 다시보기 기능을 통해 이미 중영되었거나 보지 못했던 드라마를 다시 볼 수 있게 되었다. 본 논문은 사용자의 개인화요소를 반영하여 방영되었던 많은 드라마 중 사용자들에게 가장 적합한 드라마를 추천해주는 추천 시스템을 제안한다.

### 1. 서론

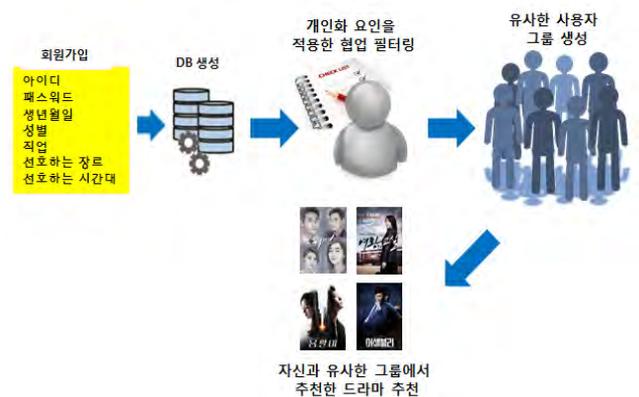
한국 국내 방송 콘텐츠의 해외수출은 <겨울연가>와 <대장금> 등 인기 드라마에 힘입어 지난 십 수 년 간 급속도로 성장했다. 이러한 ‘한류’ 열풍은 해외시장에서 한국 문화상품 및 국가 브랜드 인지도를 높이는 데 일조했음은 물론이다.[1] 이런 현상으로 한국 드라마는 유럽에서 평균 15 ~ 17개 언어로 번역되고 있으며 ‘꽃보다 남자’의 경우 한국 방송 후 3~4일 만에 20개 국어가 넘는 자막이 달리는 등 인기가 아시아에서 유럽까지 확산되고 있는 것으로 나타났다.[2] 또 최근에는 드라마에 대한 많은 인기와 관심으로 인하여 중편채널의 도입과 많은 케이블채널에서 시즌제 드라마를 제작하고 있고, 지상파 방송사는 중영된 드라마를 다시 볼 수 있는 다시보기 기능을 제공하고 있으며 일정 금액을 내고 드라마를 다시 볼 수 있는 서비스를 가진 앱이 많이 출시되면서 더욱 가깝게 사람들이 드라마를 접할 수 있는 기회가 증가하고 있는 추세이다. [3]

본 논문에서는 협업필터링의 희박성 문제 해소 방안으로 사용자가 직접 작성한 개인화 요소를 바탕으로 추천시스템의 성능을 향상시킨 드라마 추천 시스템을 설계하고 구현한다. 이러한 방법을 통해 사용자들이 원하는 드라마 추천 서비스를 보다 효율적으로 이용하도록 구현하였다.

### 2. 드라마 추천 시스템의 구성

추천 시스템을 사용하기 위해서 사용자는 회원가입을 통한 아이디, 생년월일, 성별, 선호하는 장르, 직업 및 선호하는 시간대에 대한 개인화요소로 데이터베이스를 생성한다. 관리자는 사용자에게 개인정보를 데이터베이스에 저장한 다음 사용자와 비슷한 사람들의 리스트를 구성한 다음 드라마를 추천하게 된다.

(그림 1)는 드라마 추천 절차 구성도이다.



(그림 1) 추천 절차 구성도

협업 필터링 추천 방법이란, 사용자와 성향이 비슷한 사용자들이 공통적으로 좋아하는 콘텐츠를 추천하는 방법으로 과거에 이용한 콘텐츠가 비슷하면 유사한 성향이 있다고 판단하는 근거를 토대로 이루어진다.[4] 먼저 드라마를 추천하기 앞서, 사용자들의 성향을 분류하는 기준은 (표

※ 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2014R1A1A4A01007190)

1)와 같이 선호 장르, 나이, 직업 등이다.

또한 본 논문에서는 기존 논문과 다르게 사용자의 선호하는 시간대를 입력하게 되어 있다.

번호	선호 시간대	선호 장르	직업	나이
1	오전 11시 이전	공포	학생	20세 미만
2	오후 6시~9시	판타지	주부	20세~30세
3	오후 10시 이후	코미디	직장인	31세~40세
4		액션		40세~50세
5		연애		50세~60세
6		사극		60세 이상
7		가족		
8		일반		

(표 1) 사용자들의 성향 분류 기준

### 3. 드라마 추천 시스템의 구현

개인 맞춤 추천시스템을 사용하기 위해서는 사용자의 회원가입을 필수로 하고, 회원가입 부분에서 개인화요인으로 사용할 생년월일, 성별, 국내 및 해외, 선호영화장르를 필수적으로 입력받아 회원가입이 이루어진다.[5] 회원가입 양식은 (그림 2)에서 확인 할 수 있다. 회원가입 할 때 입력한 정보는 각각 데이터베이스에 저장되고, 저장된 데이터베이스는 (그림 3)와 같다.

**회원가입**

아이디  중복체크

비밀번호  - 4~12지 이상 /  
- 특수문자 사용불가

비밀번호 확인

생년월일 2015 ▼ 1 ▼ 15 ▼

성별  남성  여성

직업  선택 ▼

선호하는 장르  선택 ▼

선호하는 시간대  선택 ▼

(그림 2) 회원가입 양식

ID	F.GENRE	BIRTH	SEX	F.TIME	JOB
13211655	6	950806	남	3	1
13251225	2	770530	남	3	3
15420205	5	910306	여	1	2
13213254	5	950531	남	2	1
13261258	7	880610	남	3	3
13212684	3	820907	남	2	1
15421063	2	841101	여	3	3
15432015	4	790129	여	1	2

(그림 3) 회원 데이터베이스

드라마를 추천하기 위해서는 드라마 정보가 들어있는 데이터베이스가 필요하다. 다음 그림들은 각 카테고리 별로 정렬된 그림이다.

(그림 4)은 선호시간대가 오후 10시 이후인 사람이 본 드라마 정보이고, (그림 5)은 사극을 좋아하는 사람이 본 드라마 정보, (그림 6)은 직장인이 본 드라마 정보를 볼 수 있다.

ID	F.TIME	Drama
15212358	3	49일
13125173	3	주몽
12133587	3	편의 전쟁
15923451	3	궁
17121231	3	허준

(그림 4) 선호시간대가 오후 10시 이후인 사람이 본 드라마 정보

ID	F.GENRE	Drama
17121231	6	동이
12671503	6	선덕여왕
13125173	6	주몽
15923451	6	이산
17215292	6	무사 백동수

(그림 5) 사극을 좋아하는 사람이 본 드라마 정보

ID	JOB	Drama
15125142	3	상속자들
13125173	3	주몽
13421332	3	연애의 발견
17251239	3	주군의 태양
13854125	3	너를 기억해

(그림 6) 직장인이 본 드라마 정보

본 논문에서는 사용자가 회원가입 할 때 입력한 개인정보에 대해서 가중치를 부여하였다. 사용자가 회원가입 할 때에 사용되는 정보 중에서 나이(10%), 성별(10%), 직업(15%), 선호하는 장르(25%) 그리고 선호하는 시간대(40%) 각각 가중치를 부여하였다. 사용자가 입력한 데이터를 바탕으로 추천받을 사용자와 유사한 성향의 사용자를 찾는다. (표 2)은 가중치의 합이 높고 가장 유사한 사용자 3명을 추출하여 드라마를 각각 하나씩 사용자에게 추천하는 Top-N 리스트 이다.

선호 시간대	선호 장르	직업	성별	나이	가중치 합	아이디
0.2	0.15	0.05	0.1	0.055	0.555	12125112
0.4	0.05	0.05	0.1	0.085	0.685	15175213
0.2	0.2	0.05	0.05	0.07	0.57	17671530
0.4	0.25	0.15	0.05	0.07	0.92	13521452
0.4	0.175	0.1	0.1	0.085	0.86	13632152
0.2	0.1	0.05	0.1	0.1	0.55	15321548
0.4	0.15	0.1	0.1	0.085	0.835	15215475
0.4	0.05	0.1	0.05	0.055	0.655	15251485
0.4	0.25	0.1	0.05	0.055	0.855	17521452
0.2	0.15	0.1	0.1	0.07	0.62	12412575
0.4	0.125	0.15	0.05	0.07	0.795	13521452
0.2	0.2	0.15	0.1	0.07	0.72	17521452
0.2	0.175	0.1	0.05	0.055	0.58	12421256
0.2	0.05	0.15	0.05	0.07	0.52	13241452
0.4	0.05	0.1	0.05	0.085	0.685	15412574
0.4	0.175	0.15	0.1	0.1	0.925	13215412
0.4	0.075	0.15	0.05	0.085	0.76	13421536
0.4	0.125	0.05	0.05	0.07	0.695	13521454
0.2	0.1	0.15	0.1	0.085	0.635	13521445
0.2	0.25	0.1	0.05	0.07	0.67	12521452
0.2	0.25	0.1	0.05	0.1	0.7	15445090
0.2	0.2	0.05	0.05	0.07	0.57	17215412
0.4	0.125	0.15	0.1	0.07	0.845	17824152
0.2	0.1	0.05	0.1	0.085	0.535	12412589
0.2	0.05	0.15	0.05	0.055	0.505	15321452
0.4	0.175	0.05	0.1	0.1	0.825	12485712
0.2	0.175	0.15	0.05	0.085	0.66	15214785
0.4	0.25	0.1	0.1	0.085	0.935	17521458
0.2	0.25	0.1	0.1	0.085	0.735	12412587
0.2	0.25	0.15	0.1	0.07	0.77	13215458
0.2	0.2	0.15	0.05	0.07	0.67	15214758

(표 2) Top-N 리스트

Top-N리스트에서 사용자에서 가중치의 값이 제일 높은 것으로 3개를 추출한다.

(그림 7)는 사용자와 성향이 비슷한 유저 13521452, 13215412, 17521458 사용자한테서 드라마 3개를 추천하는

데 가중치가 높은 상위 3명의 사용자한테 추천해주는 시스템이다.

ID	Drama
13215412	편의 전쟁
13521452	궁
17521458	49일

(그림 7) 상위 3명의 사용자 추출

이제 상위 3명의 사용자들의 드라마 정보를 추출해야 한다. 드라마 정보가 들어 있는 데이터베이스를 이용해서 추천 받고자 하는 사용자에게 드라마 정보와 사진들을 함께 보여준다. (그림 8)은 드라마 정보가 들어있는 데이터베이스 이다.

Drama	방송사	방송 기간
편의 전쟁	SBS	2007.05.16~2007.07.05
궁	MBC	2006.01.11~2006.03.30
49일	SBS	2011.03.16~2011.05.19

(그림 8) 드라마 정보 데이터베이스

(그림 9)에서 추천받은 드라마를 선택하면 아래 (그림 10)와 같이 드라마에 대한 정보를 확인 할 수 있다.

[Main](#) | [Join](#) | [Login](#)

드라마의 소개
드라마의 장르구분
드라마 추천



**편의 전쟁**

방영시간  
(오후 10~11시)



**궁**

방영시간  
(오후 10~11시)



**49일**

방영시간  
(오후 10~11시)

(그림 9) 추천 결과 화면



(그림 10) 드라마에 대한 상세 정보

#### 4. 결론

본 논문에서는 기존의 드라마 추천 시스템과 다르게 사용자의 선호하는 시간대와 장르에 가중치를 두어 구현함으로써 기존의 개인화 추천 시스템과 비교 하였을 때, 차별화하여 추천해주는 것을 보였다.

향후 연구 과제로는 본 논문에서 사용한 개인화요인 이외에 다른 요인들을 찾아 더 적합하게 드라마를 사용자에게 추천할 수 있는 것이 필요할 것으로 보인다. 현재는 사용자의 선호하는 장르와 시간대에 높은 가중치를 두어 드라마를 추천하였지만, 향후에 드라마를 추천할 때는 더 많은 개인화 요소를 추가하여 사용자가 다른 드라마를 추천 받을 수 있도록 해야 할 것이다.

#### 참고문헌

- [1] 이문행, “국내 지상파 방송드라마 해외 판매 특성 : 판매 유형, 거래 단가, 국가 간 차이를 중심으로” 수원대학교 언론정보학과 2015. 06
- [2] 홍석경, “서유럽의 동아시아 대중문화 향유를 이해하기” 보르도대학 언론정보학과 부교수 2011.5
- [3] 김영아, 박두순, “협업 필터링 기반 드라마 추천 시스템” 순천향대학교 컴퓨터소프트웨어공학과 2013.11
- [4] 김은규, “카테고리 협업 필터링 기반의 모바일 광고 추천 시스템 설계” 한양대학교 공학대학원 2012.2
- [5] 김채린, 박주현, 두추월, 박두순, “개인 리뷰를 통한 영화추천 시스템” 2014.11