

# 안드로이드에서 협업 필터링과 개인성향을 이용한 웹툰 추천 시스템

황동현\*, 임성훈\*, 박두순\*

\*순천향대학교 컴퓨터소프트웨어공학과

e-mail : pointehd@naver.com

## Webtoon recommendation system using collaborative filtering and personal propensity in Android

Dong-Hyun Hwang\*, Lim-Sung Hun\*, Doo-Soon Park\*

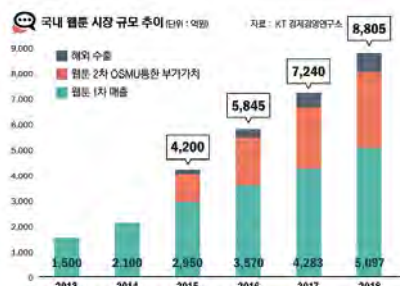
\*Dept. of Computer Software Engineering, SoonChunHyang University

### 요 약

최근 짧은 시간을 즐겁게 사용하기 위하여 어디서든 즐길 수 있는 다양한 취미가 생겨나고 있다. 그 중 하나인 웹툰은 스마트폰 환경에서 많이 사용하는데 다양한 플랫폼과 폭풍적으로 증가하는 웹툰들 중에서 자신이 원하는 웹툰을 찾기는 매우 힘들다. 본 논문에서는 안드로이드스튜디오와 R에서 협업 필터링과 개인 성향을 이용하여 개인 사용자에게 알맞은 웹툰을 추천해주는 시스템을 개발한다.

### 1. 서론

최근 바쁘게 돌아가는 현대사회 속에서 스마트폰을 이용하여 짧은 시간을 소모하고 언제 어디서든 즐길 수 있는 다양한 취미들이 떠오르고 있는데 이중 하나가 웹툰이다. 이에 따라 엄청나게 많은 종류의 웹툰이 쏟아져 나오고 있다. 대표적으로 네이버 웹툰, 다음 웹툰을 꼽을 수 있고 이외에도 다양한 플랫폼들도 생기고 있다. 네이버의 웹툰만 해도 250여개의 웹툰이 존재하며 정식 연재되지 않는 것까지 합치면 1000개 이상의 웹툰이 존재한다. 이러한 웹툰은 언제 어디서든 스마트폰을 이용하여 언제 어디서든 볼 수 있다는 장점 덕분에 계속하여 찾는 사람들이 많아지는 추세이다. 또한, 웹툰시장은 해외 수출까지 시작하면서 계속 상승세를 유지하고 있다. (그림 1)은 현재 국내 웹툰 시장 규모 추이를 나타내는 그래프이다.



(그림 1) 국내 웹툰 시장 규모 추이[1]

스마트 폰의 발달로 인해 어디서든 사용할 수 있는 앱들

\* 본 연구는 NRF-2017R1A2B1008421에 의해 지원되었음

이 사용이 많아지고 네이버 웹툰 앱이 앱사용량 14위에 들어갈 정도로 웹툰은 많은 부분을 차지하게 되었다. (그림 2)는 한국에서의 앱사용량 순위이다.



(그림 2) 한국에서의 앱사용량 순위[2]

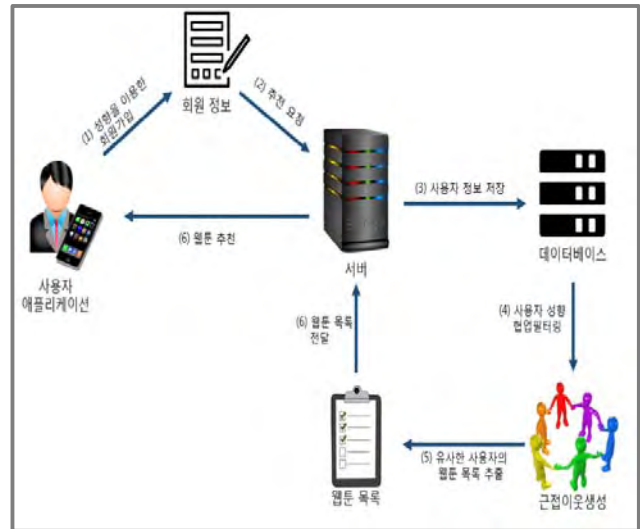
사용자의 수요에 따라 웹툰의 종류가 매우 다양해졌고 웹툰 서비스를 제공하는 사이트들도 계속해서 생겨나고 있는 추세이다. 계속되는 웹툰 시장의 확대에 의해 사용자들이 자신이 원하는 웹툰을 찾기가 힘들어졌고 사용자들의 많은 모바일 환경 선호 증가로 인하여 안드로이드 환경에서 각 개인에게 맞는 웹툰을 추천하는 것이 필요하다. 그래서 본 논문에서는 협업 필터링의 cold start 문제를 해결하기 위하여 협업 필터링과 개인 성향을 이용하여 각 개인

에 맞는 웹툰을 추천하는 시스템을 제안한다.

## 2. 웹툰 추천 시스템의 구성

상품을 추천하는데 사용되는 협업 필터링은 상품에 대하여 여러 명의 고객 추천 트랜잭션의 집합으로 구성되고 고객들 간의 유사도를 측정하여 유사한 집단을 탐색한다. 생성된 유사 집단에서 가장 유사도가 높은 고객의 추천 상품을 찾는 것이 협업 필터링이다.

협업 필터링은 새로운 사용자가 들어왔을 때 고객들 간의 유사도를 분별할 수 없어 추천의 신뢰도가 떨어지는 단점이 있다. 이것은 Cold start 문제로 사용자에게 평가 내역이 없어서 발생하는 결과이다. 이 문제를 해결하기 위해 초기에 사용자의 연령, 성별, 직업, 장르, 플랫폼 등 정보를 이용하여 Cold start 문제를 해결한다. 연령, 성별, 직업, 장르, 플랫폼 등 정보 <표 1> 과 같다.



(그림 3) 웹툰 추천 시스템 구성도

<표 1> 사용자의 성향 기준

번호	연령	성별	직업(대분류)	장르	플랫폼
1	± 5	남	관리자	에피소드	네이버
2		여	전문가 및 관련 종사자	옴니버스	다음
3			사무 종사자	스토리	레진
4			판매 종사자	일상	탑툰
5			농림어업 숙련 종사자	개그	케이툰
6			기능원 및 관련 기능종사자	판타지	카카오
7			장치·기계 조작 및 조립 종사자	액션	
8			단순 노무 종사자	드라마	
9			군인	순정	
10				감성	
11				스릴러	
12				시대극	
13				스포츠	

연령은 자신을 기준 유사한 ± 5인 사람과 비교하여 연령이 비슷할 경우 더 큰 유사도를 나타내도록 분류하였고, 직업은 통계청의 대분류를 기준을 사용하였으며, 장르는 웹툰플랫폼의 장르로 구성하였다. 웹툰을 추천 구성도는 (그림 3)과 같다.

웹툰에 대한 시나리오는 두 가지 방법을 사용하였다. 하나는 기존 사용자들이 개개인으로 있는 경우의 추천과 기존 사용자들을 클러스터링하여 그룹화 한 다음 새로운 사용자에게 대해서 추천해주는 방법이다.

(그림 3)의 추천시나리오는 사용자가 웹툰을 추천하기 전에 스마트폰을 이용하여 애플리케이션을 다운 받아야한다.

(1) 사용자는 회원가입 목록에 있는 것인 나이, 성별, 직업, 장르, 선호 플랫폼 등을 이용하여 회원가입을 한다.

(2) 회원가입에 사용된 목록들은 추천알고리즘에서 개인화 요소로 사용하기 때문에 서버로 전송한다.

(3) 서버는 사용자 정보를 DB에 저장하고 근접이웃을 생성하기 위하여 DB로부터 모든 사용자 정보를 요청한다.

(4) 모든 사용자의 개인화 요소를 이용하여 Jaccard Similarity를 이용하여 신규 사용자에게 맞는 근접이웃을 생성한다.

(5) 생성한 근접이웃 중 가장 유사도가 높은 사용자가 추천하는 웹툰을 불러온다.

(6) 생성된 웹툰 데이터를 불러와 사용자에게 3개의 웹툰에 대하여 웹툰 제목, 플랫폼, 웹툰 요일을 추천한다.

추천이 완료되고 나면 구현된 추천 시스템에 대하여 추천 알고리즘에 대한 평가가 있어야 한다. 평가는 생성된 사용자 데이터 10,000개를 대상으로 MAE(Mean Absolute Error)를 이용한다.

## 3. 웹툰 추천 시스템의 구현

웹툰 추천 시스템의 구현 중 R을 이용하여 협업 필터링을 구현한다. R은 무료소프트웨어재단의 GNU/GPL을 따르는 소프트웨어로 자유롭게 사용할 수 있으며, 다양한 통계분석 library와 강력한 그래픽 기능을 가지고 있다[3].

본 논문에서 개발하는 시스템은 Node.js로 서버를 구성하고 mongo DB로 데이터베이스를 구축한다. 이 데이터베이스를 기반으로 R에서 포스팅하여 개인화 요소, 추천 알고리즘 등의 데이터들을 다루게 된다. 또한, 안드로이드 기반시스템으로 기본 화면은 안드로이드스튜디오에 내장된 품을 이용하여 구축되도록 설계하였다.

웹툰 추천 시스템을 구현하기 위해서는 웹툰 데이터를 구축해야 하는데 웹툰 데이터는 네이버 웹툰, 다음 웹툰, 폭스툰 등 여러 플랫폼의 웹툰 데이터들로 구성하였다.

웹툰을 DB에 저장하는 방법은 Node.js의 Json을 이용하

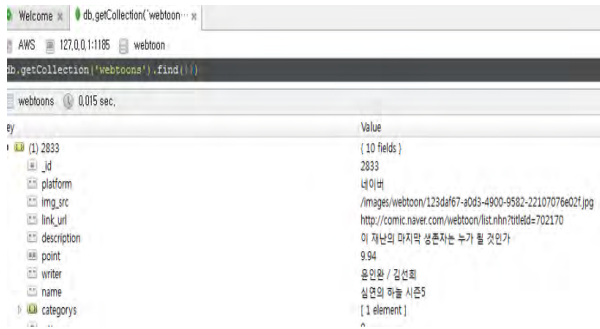
여 웹툰데이터를 전처리하여 크롤링하는 방식이다. (그림 4)는 웹툰의 기본적인 정보(플랫폼, 제목, 이미지, 작가, 요일, 회수 등)를 추출하고 있는 것이다.

```

_id: 5ab261ec3d97b32a14e56a98 }
url: { platform: '네이버',
       url: '/webtoon/list.nhn?titleId=22873',
       category: 2,
       _id: 5ab261ec3d97b32a14e56a91 }
url: { platform: '네이버',
       url: '/webtoon/list.nhn?titleId=671421',
       category: 2,
       _id: 5ab261ec3d97b32a14e56a92 }
url: { platform: '네이버',
       url: '/webtoon/list.nhn?titleId=642598',
       category: 2,
       _id: 5ab261ec3d97b32a14e56a93 }
url: { platform: '네이버',
       url: '/webtoon/list.nhn?titleId=24777',
       category: 2,
       _id: 5ab261ec3d97b32a14e56a94 }
url: { platform: '네이버',
       url: '/webtoon/list.nhn?titleId=675556',
       category: 2,
       _id: 5ab261ec3d97b32a14e56a95 }
url: { platform: '네이버',
       url: '/webtoon/list.nhn?titleId=678151',
       category: 2,
       _id: 5ab261ec3d97b32a14e56a96 }
    
```

(그림 4) 웹툰 DB를 생성하는 과정

(그림 5)는 웹툰의 정보가 저장되어있는 Mongo DB의 화면이다. DB에는 플랫폼의 이름, 이미지, 제목, 평점, 작가, 줄거리가 포함되어있다.



(그림 5) 데이터베이스 웹툰 저장

신규 사용자가 추천을 요청할 경우 사용자의 개인요소를 R을 이용하여 불러오며 결과의 일부는 (그림 6)과 같다.

user.number	user.age	user.gender	user.job	user.genre	user.platform
1	1	25	1	4	3
2	2	24	2	3	4
3	3	26	1	2	11
4	4	38	1	3	13
5	5	40	1	4	2
6	6	15	2	5	3
7	7	22	1	6	4
8	8	33	1	7	5
9	9	35	2	1	8
10	10	47	1	2	9
11	11	61	1	1	10
12	12	40	1	2	11

(그림 6) R Studio 사용자 개인요소

R에서 지원하는 'recommenderlab'를 이용하여 Jaccard Similarity를 이용하여 사용자들 간에 유사도를 측정하여

근접이웃 리스트를 생성한다. Jaccard Similarity는 식 (1)과 같다.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad \text{--- (1)}$$

Jaccard Similarity는 A와 B 사용자 간에 유사한 것을 모든 경우로 나누는 것이다 가장 유사할 경우 1의 값을 가진다.

나이 23, 남성, 직업은 기능원 및 관련 기능종사자, 장르는 에피소드, 플랫폼은 네이버인 943번 사용자에게 대해서 사용자간 유사도를 비교한 일부 값은 (그림 7)과 같다.

u.number	age	u.gender	u.job	u.genre	u.platform	u.similarity
943	23	1	6	1	1	1.00
866	19	1	6	1	1	0.91
134	24	1	6	1	1	0.70
423	28	1	6	1	1	0.67
486	25	1	5	3	3	0.50
57	24	1	3	3	3	0.47
135	27	1	11	2	2	0.45
762	23	2	13	2	2	0.40
895	23	2	6	3	3	0.40

(그림 7) 943번 사용자 유사도 리스트

(그림 7)의 유사도 리스트에서 유사도가 가장 높은 사용자 top-3 리스트는 (그림 8)과 같다. top-3 리스트에 있는 사용자들로부터 추천된 웹툰은 신도림, 이태원클라스, TEN이다.

u.number	u.similarity	w.id	webtoon
866	0.91	3889	신도림
134	0.91	4864	이태원클라스
423	0.90	1924	TEN

(그림 8) 943번 사용자의 웹툰 추천 결과

다음으로는 사용자 개개인에 대해서 유사도를 측정하는 것이 아닌 사용자들에 대해서 먼저 clustering을 해서 그룹을 만들고 새로운 사용자가 들어올 때 새로운 사용자와 그룹들에 대한 유사도를 비교한다. (그림 9)는 943번 사용자에게 대해서 그룹화 되어있는 집단과 유사도를 판단하는 것이다.

number	age	gender	job	genre	platform	similarity
943	23.0	1	6	1	1	1.00
20G	24.6	1	6	1	1	0.96
4G	24.0	1	6	6	1	0.92
48G	21.4	1	6	1	1	0.87
15G	23.0	1	6	1	3	0.80
41G	22.0	2	6	1	1	0.78
16G	27.0	1	3	1	1	0.71

(그림 9) 그룹화 유사도 리스트

가장 유사도가 높은 집단에 새로운 사용자를 넣고 집단에 있는 사용자들이 추천하는 웹툰 중 가장 많이 선택되고 별점이 높은 것을 새로운 사용자에게 추천해준다. (그림 10)은 사용자가 들어간 집단의 웹툰 추천 목록이다.

number	age	gender	job	genre	platform	group	wid	webtoon	score
1	349	24	1	6 1	1 20G	20G	1894	신의탑	4.5
2	179	22	2	6 1	2 20G	20G	1894	신의탑	4.0
3	866	24	1	4 1	2 20G	20G	2138	신도림	4.5
4	678	23	1	6 1	1 20G	20G	2571	스위트홈	5.0
5	221	23	1	6 1	1 20G	20G	1534	TEN	4.0
6	943	23	1	6 1	1 20G	20G	0		0.0

(그림 10) 집단에서 추천하는 웹툰 목록

여기서 추천된 웹툰은 신의탑, 신도림, 스위트홈이다. 943번 사용자가 안드로이드 스튜디오에서 지원하는 애플레이터를 이용하여 애플리케이션을 실행하여 결과를 받으면 (그림 11)와 같다.



(그림 11) 애플리케이션 추천 결과

추천 되는 두 가지 방법에 대해서 MAE평가 방법을 사용하여 평가한다. MAE(Mean Absolute Error)는 식 (2)와 같다.

$$MAE = \frac{\sum_{i=1}^q |실제고객평가치_i - 예측된평가치_i|}{q} \quad \text{--- (2)}$$

단, q는 사용자가 평가를 한 아이템 개수이고 i는 아이템의 평가 점수이다. 이것을 이용하여 실제 평가치와 예측 평가치의 합을 아이템의 개수로 나누어 오차 정도를 나타낼 수 있다. 사용자 데이터 10,000개를 대상으로 6,000개는

학습데이터, 4,000개는 테스트데이터로 구성하여 테스트 데이터를 가지고 (식 2)를 이용하여 MAE를 평가한 결과는 <표 2>와 같다.

<표 2> MAE 결과 편차

Person Compare MAE	Group Compare MAE
1.2741	1.1683

<표 2>의 결과 Person Compare MAE는 사용자의 개인 요소를 모든 사람과 비교하고 예측 평가치와 실제 고객의 평가 차의 평균이고 Group Compare MAE는 기존에 사용자들 성향으로 그룹을 만들고 새로운 사용자들 그룹에 넣고 그룹에서 추천하는 웹툰을 받아 예측 평가치와 실제 고객의 평가 차의 평균이다.

MAE의 결과를 보면 Group Compare MAE가 0.11만큼 낮은 것을 볼 수 있다. 모든 사용자에 대하여 비교를 하고 가장 높은 사용자의 웹툰을 추천 받는 것 보다 그룹화 되어 있는 곳에 들어가 유사한 사용자들이 추천해주는 여러개의 웹툰 데이터가 더 정확하다고 볼 수 있다. 그룹화를 통하여 더 정확하게 cold start 문제를 웹툰의 평가 없이 성향을 이용하여 해결 할 수 있다.

#### 4. 결론

본 논문에서 어디서든 사용할 수 있는 스마트폰의 발달로 인하여 짧은 시간을 통하여 즐길 수 있는 취미들 중 하나인 웹툰을 사용자에게 맞게 추천하여 준다. 또한 협업필터링의 Cold start 문제를 해결하기 위하여 개인 성향을 Jaccard Similarity를 이용하여 근접이웃을 생성하여 초기 웹툰을 추천해준다. 향후 연구 과제로 사용자의 데이터를 직접 만들고 평가하는 것이 아닌 실제 사용자들의 평가 데이터를 수집하여야 할 것이고 더 많은 사용자들의 데이터를 이용하여 평가하여야 한다.

#### 참고문헌

- [1] 최진기, “일상 소재들로 만든 5800억 시장... 지금은 웹툰시대”, 가천대신문 2016.03
- [2] 김준범 “한국·일본 앱 MAU랭킹 비교!-가장많은 사람들이 사용하는 앱은?”, AppApp 2016.05
- [3] 고영준, 박용민, 김진석 “WebER:R을 이용한 웹 기반의 교육용 통계 분석 시스템 구현”, 한국통계학회논문집, 19권, 2호, pp. 257-266, 2012