

클라우드 컴퓨팅 환경에서 협업필터링과 개인 성향을 이용한 개인화 소셜 추천 시스템

장태훈*, 김한이*, 박두순*
 *순천향대학교 컴퓨터소프트웨어공학과
 e-mail : wkdxogns23@naver.com

A Personalized Novel Recommendation System based on Collaborative Filtering and Personal Propensity in Cloud Computing Environment

Tae-Hoon Jang, Han-Yi Kim, Doo-Soon Park
 *Dept. of Computer Software Engineering, SoonchungHyang University

요 약

최근 바쁜 일상 속에서 개인의 삶의 질과 활력을 높이기 위해 여가활동에 대한 관심이 증가하고 있고 그 중에서 독서는 꾸준한 사랑을 받고 있는 여가 활동이다. 그 중 소설의 출판량은 다른 타 장르에 비해 가히 압도적이다. 하지만 소설은 개인의 취향에 영향을 많이 받는다는 특징이 있어 사용자에게 적합한 소설을 추천하기란 기존의 시스템으로는 한계가 있다. 따라서 본 논문에서는 클라우드 컴퓨팅 시스템인 AWS(Amazon Web Service)를 이용하며 사용자의 개인 성향과 협업 필터링 방법을 이용하여 각각의 개인 성향에 적합한 소설을 추천하는 시스템을 제안한다.

1. 서론

최근 바쁜 일상 속에서 개인의 삶의 질과 활력을 높이기 위해 여가활동에 대한 관심이 증가하고 있다. 많은 여가활동 중에서 독서는 오래전부터 수많은 사람들에게 꾸준한 사랑을 받고 있는 취미이다.

통신기술의 발달과 스마트기기 사용자의 증가로 모바일 서적의 접근이 쉬워졌고 작가들이 선택할 수 있는 창작 채널과 창작 플랫폼이 갈수록 많아졌다. 그에 따라 자연스럽게 수요와 공급이 증가하게 되면서 도서 시장이 확대되었다. 그 중 소설은 다른 도서 장르에 비해 기존 도서, 신간도서의 출판량이 가장 높은 비율을 차지하고 있다. (그림 1)은 2013년과 2014년 문학도서 장르별 신간도서 현황 이다[1].

(표 1) 2013-2014년 문학 도서 장르별 신간도서 현황

구분	시	희곡	소설	수필	평론	기타	계
2014년	1,399	485	5,647	541	16	2,583	10,671
2013년	1,122	55	5,327	495	16	2,281	9,296

기존의 도서 구매사이트에서 사용되고 있는 추천 시스템들은 단지 많이 구매된 베스트 셀러들을 추천하고 있다. 다른 도서들에 비해 소설은 '흥미'를 위해 접근을 많이 하는 특징이 있고 개인의 취향에 영향을 많이 받는다.

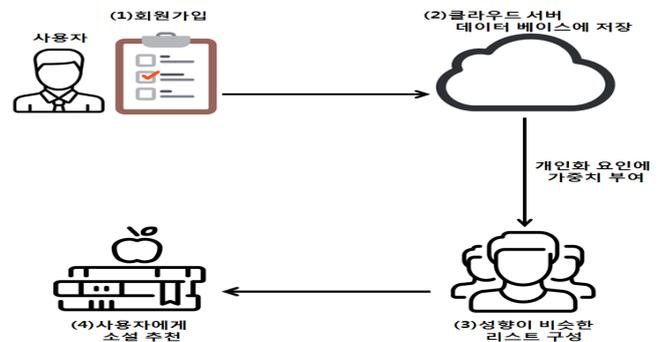
이와 같은 방법은 사용자의 기호에 맞추기에 매우 부족하다.

본 논문에서는 개인화 요인에 가중치를 부여한 것을 바탕으로 협업 필터링을 사용하여 사용자 유사성 판별, 후 사용자의 성향에 적합한 소설을 추천하는 시스템을 제안한다.

2. 소설 추천 시스템의 구성

추천 시스템에는 여러 가지 기법이 사용되고 있으나, 본 논문에서는 협업필터링 기법을 사용하여 클라우드 컴퓨팅에서 시스템을 구현하였다.

소설 추천 시스템의 시나리오는 (그림 1)과 같다.



(그림 1) 추천 시스템 시나리오

추천 시스템은 (1)협업필터링을 사용하기 위한 데이터를 회원가입시 사용자로부터 개인화 요인으로 사용하게 될 나이, 성별, 직업, 선호하는 국내/국외 소설 그리고 선호하는 장르를 입력받고 AWS 클라우드 서버 데이터베이스에 저장한다. (2)각 개인화요인에 가중치를 부여한다. 연령대(20%), 성별(10%), 직업(25%), 국내/국외 소설(5%), 선호 장르(40%)로 부여한다. (3)그 후 가중치에 따라 사용자와 비슷한 성향의 다른 사용자를 찾아 Top-n리스트를 구성한다. (4)Top-n리스트를 기반으로 사용자에게 소설을 추천하게 된다.

3. 소설 추천 시스템의 구현

본 논문에서는 아마존 웹 서비스(AWS)의 대표적인 IaaS 서비스인, EC2 인스턴스를 사용하여 개발 환경을 구축하였다. AWS EC2 인스턴스는 Linux 콘솔 환경으로 구성되어 있어, 클라우드 환경에 가장 빠르게 접속시킬 수 있는 Ruby on Rails 프레임워크를 사용하여 개발하였다 [2].

(그림 2)는 가중치를 부여하지 않고 추천된 소설이다. 개인화 요소의 일치 여부만을 확인하여 추출한 Top-N리스트의 상위 10명을 추출한 리스트이다.

	연령	성별	직업	선호 장르	국내/국외 도서	총합	소설
User_id 104	20대 초반	여자	사무직	판타지	국내	1.0	싸울아비 룬, 달빛 조각사, 무림드래곤
User_id 654	10대 후반	남자	사무직	판타지	국외	0	헝거게임, 트와일라잇, 해리포터
User_id 665	40대 초반	여자	서비스업	동화	국내	0	강아지 동물, 작은 손톱
User_id 721	30대 초반	여자	군인	로맨스	국내	0	도레미파울라시도, 그놈은 멋있었다
User_id 85	20대 초반	여자	계조작 및	현대	국외	0	내 심장을 쫓아, 연둑이
User_id 365	40대 후반	남자	판매직	동화	국내	0	태국우준, 군림하, 비뢰도
User_id 042	30대 초반	남자	서비스업	판타지	국외	0	아이 엠 넘버 포, 윌 베디스
User_id 421	50대 후반	여자	서비스업	로맨스	국외	0	개인의 취향, 어떤날 그녀들이, 사랑하는 사람이 생겼습니다.
User_id 048	30대 후반	남자	관리자	현대	국내	0	악마는 법정에 서지 않는다, 황태자의 남자사건
User_id 99	40대 초반	남자	사무직	현대	국외	0	국화꽃 향기, 동대지기, 시골의사의 아름다운 동행
User_id 288	60대 초반	여자	계조작 및	추리	국외	0	카지노, 환상의 여인, 얼음나무 숲

(그림 2) 가중치를 부여하지 않고 추천된 소설

(그림 3)은 가중치를 부여해 개인화요인을 사용하여 추천한 Top-N 리스트이다.

	연령	성별	직업	선호 장르	국내/국외 도서	총합	소설
User_id 104	0.2	0.1	0.25	0.4	0.05	1.0	싸울아비 룬, 달빛 조각사, 무림드래곤
User_id 654	0.19	0.05	0.25	0.4	0.025	0.915	헝거게임, 트와일라잇, 해리포터
User_id 042	0.18	0.05	0.13	0.40	0.025	0.785	아이 엠 넘버 포, 윌 베디스
User_id 421	0.13	0.1	0.13	0.34	0.025	0.725	개인의 취향, 어떤날 그녀들이, 사랑하는 사람이 생겼습니다.
User_id 665	0.16	0.1	0.13	0.28	0.05	0.72	강아지 동물, 작은 손톱
User_id 048	0.17	0.05	0.22	0.22	0.05	0.71	악마는 법정에 서지 않는다, 황태자의 남자사건
User_id 82	0.18	0.1	0.04	0.34	0.05	0.71	도레미파울라시도, 그놈은 멋있었다
User_id 82	0.13	0.05	0.16	0.32	0.05	0.71	역해웅주, 핵을 품은 달, 남한산성
User_id 15	0.19	0.05	0.22	0.22	0.025	0.705	그리고 아무도 없었다, V의 비극
User_id 99	0.15	0.05	0.25	0.22	0.025	0.695	국화꽃 향기, 동대지기, 시골의사의 아름다운 동행
User_id 365	0.15	0.05	0.16	0.28	0.05	0.69	태국우준, 군림하, 비뢰도

(그림 3) 개인화 요인에 가중치를 부여해 추천된 소설

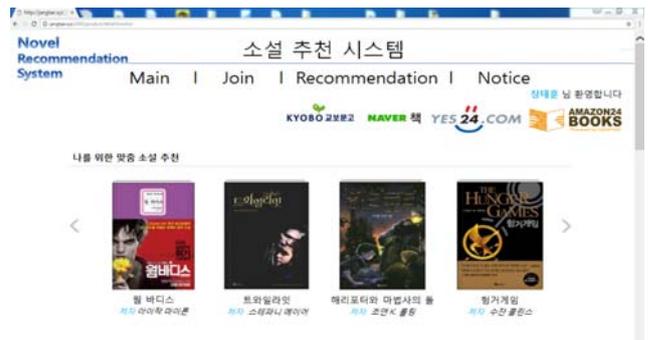
(그림 2)와 (그림 3)을 분석해본 결과 가중치의 유무에 따라 사용자에게 추천된 소설이 다르게 나온 것을 확인할 수 있다. (그림2)에서는 소설들이 장르에 관계없이 추천된 것을 볼 수 있다. 그에 비해 개인화 요인의 선호 장르에 가중치를 40%주었을 때 (그림 3)의 상위 3명 모두 선호 장르가 판타지인 것을 볼 수 있다. 이를 통해 선호 장르가

가장 큰 영향을 받아 이를 바탕으로 소설이 추천되었다. 이와 같은 방법으로 개인화 요인의 가중치를 변경해가며 분석해 본 결과로 협업 필터링 방법에 연령대(20%), 성별(10%), 직업(25%), 국내/국외 소설(5%), 선호 장르(40%)로 가중치를 부여하였다. Top-N리스트 중 2명을 추출하여 추천하는 소설은 (그림 4)와 같다.

	연령	성별	직업	선호 장르	국내/국외 도서	총합	소설
User_id 104	0.2	0.1	0.25	0.4	0.05	1.0	싸울아비 룬, 달빛 조각사, 무림드래곤
User_id 654	0.19	0.05	0.25	0.4	0.025	0.915	헝거게임, 트와일라잇, 해리포터
User_id 042	0.18	0.05	0.13	0.40	0.025	0.785	아이 엠 넘버 포, 윌 베디스

(그림 4) 개인화 요인에 가중치를 부여해 추천된 Top-N 리스트

사용자와 가장 유사한 2명의 소설 추천 리스트를 보면 5권의 소설이 존재한다. 본 논문에서는 가중치를 가장 많이 주었던 선호 장르가 비슷한 소설 순서로 추천한다. 또한 (그림 5)는 추천 결과이고 추천된 소설을 바로 구매하고 싶거나, 정보를 쉽게 알아 볼 수 있게 상단에 e-book을 구매할 수 있는 사이트를 링크함으로써 사용자가 편리하게 소설을 접할 수 있게 구성하였다.



(그림 5) 추천 결과

4. 결론

본 논문은 사용자의 성향에 부합하는 적절한 소설을 보다 편리하고 효과적으로 추천하기 위해 클라우드 컴퓨팅 시스템에서 개인화 요인과 협업 필터링을 기반으로 구현한 추천 프로그램 시스템이다.

향후 연구 과제는 소설을 다양하게 추천하기 앞서 소설의 분류가 정확하게 이루어져야 할 것이고, 추천의 지표가 되는 개인화 요인을 변화시켜 추천하거나 개인화 요인의 수를 추가하여 보다 정확하게 소설을 추천하여 본 논문에서 제안한 것 보다 한층 더 발전된 방법으로 사용자의 성향에 부합하는 소설을 추천할 수 있도록 할 것이다.

참고문헌

- [1] 김영아, 박두순, “협업 필터링 기반 드라마 추천 시스템”, 한국정보처리학회 추계 학술발표대회 논문집, 제2-권 제2호, 제주한라대학교, pp. 1137-1138, 2013.11
- [2] 이건호, 박두순, “클라우드 컴퓨팅에서 구축한 협업 필터링 기반 웹툰 추천 시스템”, 한국정보처리학회 춘계 학술대회 발표 논문집, 동국대학교, pp. 451-454, 2016.05