

개인화 요인과 협업필터링 기반의 도서 추천 시스템

정연우*, 박두순*
*순천향대학교 컴퓨터소프트웨어공학과
e-mail : yeonwoo93@naver.com

A Book Recommendation System based Collaborative Filtering and Personal Elements

Yeon-Woo Jeong, Doo-Soon Park
Dept. of Computer Software Engineering, Soonchunghyang University

요 약

최근, 수많은 종류의 도서가 출판되고 있다. 또한 도서의 분야와 장르, 종류가 다양해지고 그 양 역시도 방대해지고 있다. 이러한 상황에서 사용자에게 적절한 도서를 고르기란 어려운 일이다. 본 논문에서는 보다 편리하고 적절한 도서 선택을 위해 도서추천시스템을 제안한다. 사용자의 나이와 성별, 국내/외도서, 선호 장르에 가중치를 부여하고 협업필터링을 사용하는 추천 시스템을 제안한다.

1. 서론

정보화 시대가 도래 하면서 정보의 중요성이 부각되고 있다. 그럼에도 책은 여전히 정보를 얻을 수 있는 가장 쉬운 방법 중에 하나이다. 그에 따라 오늘날에는 수많은 종류의 도서가 출판되고 있으며, 도서의 분야와 장르, 종류가 날로 다양해지고 양도 방대해지고 있다. 새로운 정보와 지식에 대해 도서를 통해 쉽게 얻을 수 있지만, 도서의 종류와 분야 등이 다양해지고 방대해짐에 따라 어떤 도서를 선택해야 할지 막막한 것 또한 사실이다[1].

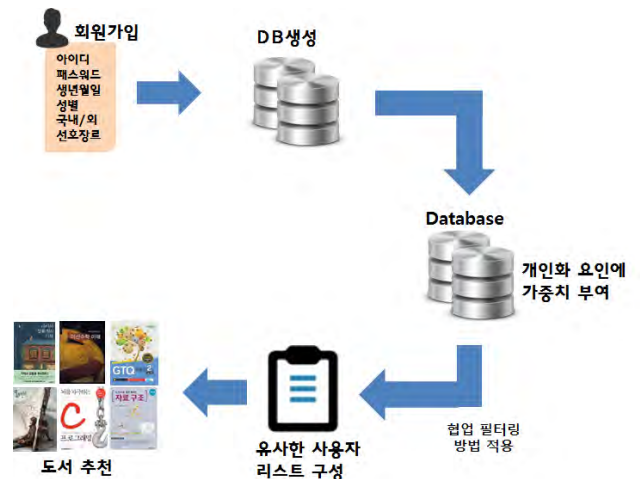
이러한 도서의 선택에 있어서 도움을 주기 위해서 아마존(www.amazon.com) 이나 yes24(www.yes24.com) 와 같은 인터넷 도서 판매 사이트에서 도서 추천을 해주는 시스템이 있지만, 해당 추천 시스템이 사용자들의 기호와 상황에 부합하기란 어려운 것 역시 사실이다.

본 논문에서는 사용자 도서 선택 편의를 위한 보다 적합한 도서 추천 시스템을 제안한다. 사용자들이 회원가입 시 입력한 개인화 요인을 데이터베이스로 저장하여 이를 바탕으로 협업 필터링을 사용하여 이용자에게 보다 정확한 추천을 할 수 있게 한다.

2. 도서 추천시스템의 구성

본 논문에서 구현한 도서 추천 시스템은 (그림 1) 과 같다

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



(그림 1) 추천 절차 구성도

본 논문에서 제안하는 추천 시스템을 사용하기 위해서는 사용자로부터 회원가입을 통해 개인화 요인인 생년월일, 성별, 국내/외 도서의 종류, 선호 장르를 입력받는다. 입력받은 개인화 요인들은 데이터베이스에 저장한 후에, 협업필터링을 이용하여 사용자와 유사한 사용자 리스트를 구성하게 된다. 이러한 유사 사용자에게 적절한 도서를 추천하게 된다.

협업 필터링이란, 사용자들의 선호도와 관심 표현을 바탕으로 선호도, 관심도가 비슷한 사용자들 식별해 내는 방법으로 과거에 이용한 콘텐츠가 비슷하다면 유사한 성향을 가지고 있다고 판단하는 근거를 토대로 이루어진다[2].

협업 필터링을 사용하기 위해 사용자의 개인화 요인 항목별로 비교를 하게 된다. 따라서 각 항목별로 가중치를 부여 하여 계산된 값이 1이 넘지 않는 한도 내에서 1에 가장 근접한 사용자를 간추려 리스트를 구성하게 된다. 이러한 리스트를 통해 사용자에게 도서를 추천하게 된다.

본 논문에서 도서의 장르를 분류 하였는데, (그림2)와 같이 대표적인 장르 6가지로 간추려서 분류 하였다.

SF/판타지	공포/스릴러	로맨스
무협	전쟁/역사	추리

(그림 2) 도서의 선호 장르 분류

3. 도서 추천 시스템의 구현

본 논문에서 구현한 도서 추천 시스템을 이용하기 위해선 회원가입이 필요하다. 개인화 요인으로 사용될 생년월일, 성별, 선호 장르, 국내/외 도서의 종류를 필수적으로 입력받아 회원가입이 이루어지게 된다. 회원가입의 양식은 (그림 3)와 같다.

(그림 3) 회원가입 양식

사용자가 입력한 정보는 회원관리 테이블에 저장되고, 회원관리 테이블은 선호 장르, 국내/외 도서의 종류, 생년월일, 성별, ID 순으로 저장된다. 회원관리 데이터베이스는 (그림 4)와 같다.

ID	GENRE	BIRTH	SEX	NATION
124002	3	19930307	남	1
134050	1	19890822	남	1
114332	2	19960320	여	2
115623	5	19900314	남	1
136452	4	19900423	남	2
123595	6	19920923	남	2
154087	3	19880518	여	1
142505	4	19861211	여	1

(그림 4) 회원 관리 데이터베이스

도서를 추천하기 위해서는 사용자들의 데이터베이스뿐만 아니라 추천할 도서들의 데이터베이스 역시도 필요하다. 다음은 (그림 5)는 공포/스릴러 장르를 선호하는 사용자들이 본 도서 정보를 추출한 데이터베이스 이고, (그림 6)은 국내 도서를 선호하는 사용자들이 본 도서 정보를 추출한 데이터베이스 이며, (그림 7)은 남성들이 선택한 도서 정보를 추출한 데이터베이스이다.

ID	GENRE	BOOK	SEX	NATION
135682	2	있어 없어?	남	1
134600	2	거울 속	남	1
125860	2	종말일기	여	2
112330	2	나쁜 책	남	1
104038	2	저단	남	2
104250	2	푸른 불꽃	남	2

(그림 5) 공포/스릴러 장르를 선호하는 사용자들이 본 도서 정보

ID	NATION	BOOK	SEX	GENRE
124050	1	글자전쟁	남	1
124044	1	가짓말	남	4
104003	1	선의 법칙	여	2
093022	1	트렁크	남	3
154098	1	주름	남	3
159876	1	국경시장	남	6

(그림 6) 국내도서를 선호하는 사용자들이 본 도서 정보

(그림 7) 남성 사용자들이 본 도서 정보

본 논문에서는 사용자가 회원가입 할 때 입력한 개인화 요인을 각각 성별(10%), 나이(30%), 국내/외 도서(20%), 선호 장르(40%)로 가중치를 부여하였다. 이를 통해 성향

ID	SEX	BOOK	GENRE	NATION
094033	남	마법의 해변	1	1
153829	남	연애 독본	6	1
146785	남	파병 부대	4	2
113398	남	아름	2	1
124054	남	거울 속	3	2
124432	남	모모	5	2

이 비슷한 사용자들 찾는다. 찾은 사용자들을 TOP-N 리스트는 (그림 8)과 같다. TOP-N 리스트를 정렬하여 추천 받을 사용자와 가장 유사한 사용자 5명을 추출하고, 그 사용자와 얻은 도서들을 정렬하여 상위 3개의 도서는 (그림 9)와 같다.

선호장르	국내/외	나이	성별	가중치 합
0.4	0.3	0.2	0.1	1
0.38	0.3	0.2	0.05	0.93
0.18	0.3	0.036	0.05	0.566
0.35	0.3	0.2	0.1	0.95
0.24	0.15	0.02	0.05	0.46
0.24	0.15	0.1	0.1	0.59
0.27	0.15	0.148	0.1	0.668
0.27	0.3	0.125	0.05	0.745
0.18	0.3	0.06	0.05	0.59
0.27	0.15	0.2	0.1	0.72
0.3	0.15	0.074	0.1	0.624
0.3	0.15	0.1	0.05	0.6
0.27	0.3	0.2	0.1	0.87
0.21	0.15	0.095	0.05	0.505
0.21	0.3	0.1	0.05	0.66
0.24	0.3	0.2	0.05	0.79
0.21	0.15	0.042	0.1	0.502
0.24	0.3	0.2	0.05	0.79
0.24	0.15	0.025	0.1	0.515
0.3	0.3	0.103	0.1	0.803
0.38	0.15	0.03	0.1	0.66
0.3	0.15	0.1	0.05	0.6
0.21	0.15	0.08	0.1	0.54
0.18	0.3	0.015	0.05	0.545
0.21	0.15	0.048	0.05	0.458
0.18	0.15	0.005	0.1	0.435
0.35	0.3	0.2	0.1	0.95
0.21	0.3	0.054	0.05	0.614
0.21	0.3	0.042	0.1	0.652
0.21	0.3	0.01	0.05	0.57
0.31	0.3	0.2	0.1	0.91
0.21	0.3	0.06	0.1	0.67
0.27	0.3	0.088	0.1	0.758
0.18	0.15	0.136	0.5	0.966
0.18	0.15	0.088	0.05	0.468

(그림 8) TOP-N 리스트

선호장르	국내/외	나이	성별	가중치 합
0.4	0.3	0.2	0.1	1
0.38	0.3	0.2	0.1	0.98
0.35	0.3	0.2	0.1	0.95
0.35	0.3	0.185	0.1	0.935

(그림 9) TOP-N 리스트에서 추출한 상위 3개 도서

이 추출한 모습을 사용자에게 (그림 10)과 같은 화면을 통해 알려주며, 도서를 선택하여 클릭하게 되면, (그림 11)과 같은 화면이 출력되면서, 해당 도서의 저자, 출판사, 장르를 알려주고, 책에 대한 소개를 보여줌으로서 사용자가 추천받은 도서에 대한 정보를 알 수 있도록 구성하였다.

(그림 10) 상위 3개의 도서를 추천하는 화면

(그림 10) 추천도서를 클릭했을 때의 화면

4. 결론

본 논문은 수많은 도서 중에서 사용자가 원하는 적절한 도서를 보다 편리하고 효과적으로 추천하기 위해 협업 필터링을 기반으로 추천 프로그램을 구현하였다.

향후 연구과제로는 본 논문에서 분류한 장르뿐만이 아니라 도서에 대한 분야에 대한 선택을 할 수 있게끔 하여 사용자가 보다 폭넓은 분야와 장르의 도서를 선택할 수 있게 하거나, 개인화 요인에 대한 가중치와 협업필터링 만이 아니라 다른 방법을 이용하여 추천 할 수 있는 추천 시스템을 구현하는 것이 필요한 것으로 보인다.

참고문헌

- [1] 임찬식, 이원재, 이하나, 이세화, 이상준, “MapReduce Model에 기반한 도서 추천 시스템의 설계 및 구현” 숭실대학교 컴퓨터학부, 광운대학교 컴퓨터소프트웨어학과 2010.6
- [2] 김영아, 박두순, “협업 필터링 기반 드라마 추천 시스템” 순천향대학교 컴퓨터소프트웨어공학과 2013.11