

대출 기록에 기초한 대학 도서관 도서 개인화 추천시스템 개발 및 평가에 관한 연구

A Study on the Development and Evaluation of Personalized Book Recommendation Systems in University Libraries Based on Individual Loan Records

홍연경 (Yeonkyoung Hong)*, 전서영 (Seoyoung Jeon)**
최재영 (Jaeyoung Choi)***, 양희운 (Heeyoon Yang)****
한채은 (Chaeun Han)*****, 주영준 (Yongjun Zhu)*****

초 록

본 연구는 대학 도서관 사용 증진을 위하여 개인별 맞춤 도서 추천시스템을 구축하는 것을 목적으로 한다. 특히 사용자의 아이টে에 대한 선호도가 존재하는 다수의 추천시스템과는 달리, 선호도가 존재하지 않을 때에 도서 추천이 가능하도록 하는 방안인 도서관 이용자의 도서 대출 목록과 성향을 활용하여 평가지표를 생성하는 방법을 제안하고자 한다. 이용자가 아직 읽지 않은 책에 대한 예상 선호도를 산출하는 방식으로 도서를 추천하는 행렬 분해 방법인 Singular Value Decomposition(SVD)과 Stochastic Gradient Descent(SGD) 알고리즘을 활용한 모델을 구축했다. 더불어 유사도가 높은 이용자 그룹 내의 도서 대출 목록을 참조하여 추천하는 사용자 기반 협업 필터링 알고리즘을 활용해 모델을 구현했다. 최종적으로 평가지표를 활용한 세 가지 모델에 대하여 사용자 평가를 진행했다. 각각의 모델이 제시한 개인별 맞춤 도서 다섯 권의 목록을 해당 대출자에게 제공하고, 추천 도서에 대한 만족/불만족 여부를 이진화 점수화하여 모델에 대한 평가를 진행했다.

ABSTRACT

The purpose of this study is to propose a personalized book recommendation system to promote the use of university libraries. In particular, unlike many recommended services that are based on existing users' preferences, this study proposes a method that derive evaluation metrics using individual users' book rental history and tendencies, which can be an effective alternative when users' preferences are not available. This study suggests models using two matrix decomposition methods: Singular Value Decomposition(SVD) and Stochastic Gradient Descent(SGD) that recommend books to users in a way that yields an expected preference score for books that have not yet been read by them. In addition, the model was implemented using a user-based collaborative filtering algorithm by referring to book rental history of other users that have high similarities with the target user. Finally, user evaluation was conducted for the three models using the derived evaluation metrics. Each of the three models recommended five books to users who can either accept or reject the recommendations as the way to evaluate the models.

키워드: 대학 도서관, 추천시스템, 개인화
university library, recommendation system, book recommendation

-
- * 성균관대학교 문헌정보학과 학사과정(hyk0829@g.skku.edu) (제1저자)
 - ** 성균관대학교 문헌정보학과 학사과정(cjaktks98@g.skku.edu) (공동저자)
 - *** 성균관대학교 문헌정보학과 학사과정(cjengy@g.skku.edu) (공동저자)
 - **** 성균관대학교 문헌정보학과 학사과정(chri0220@g.skku.edu) (공동저자)
 - ***** 성균관대학교 문헌정보학과 학사과정(hancece@g.skku.edu) (공동저자)
 - ***** 성균관대학교 문헌정보학과 조교수(yzhu@skku.edu) (교신저자)

- 논문접수일자: 2021년 5월 17일 ■ 최초심사일자: 2021년 6월 4일 ■ 게재확정일자: 2021년 6월 12일
- 정보관리학회지, 38(2), 113-127, 2021. <http://dx.doi.org/10.3743/KOSIM.2021.38.2.113>

* Copyright © 2021 Korean Society for Information Management
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

정보 통신 기술의 발달로 수많은 제품과 콘텐츠들이 생산되고 있다. 이에 결정권자인 소비자를 돕는 추천시스템에 대한 관심이 점차 대두되고 있고 이미 여러 분야에서 광범위하게 활용되고 있다. 동영상 플랫폼인 넷플릭스(Netflix)의 경우, 가입 시 관심 있는 장르를 선택하게 한다. 이후 사용자가 시청한 영상 목록과 그 영상들에 대한 해당 사용자의 평가를 반영하여 시청자 개인의 취향에 맞는 영상을 추천하는 방식으로 서비스를 제공하고 있다. 세계적인 온라인 쇼핑몰 아마존(Amazon)은 사용자의 이전 구매 기록을 활용하여 관심 가질 만한 아이템들을 추천하는 방식을 활용하고 있다. 그들만의 자체 알고리즘을 개발하였으며 이것이 아이템 기반 협업 필터링에 해당한다(Linden, Smith, & York, 2003). 아마존은 이러한 알고리즘을 활용하여 사용자들이 남긴 제품에 대한 정보와 사용자 본인에 대한 정보를 활용하여 아이템 추천이 이루어지게 한다.

대학 도서관에는 이미 수많은 책이 존재하고 매년 새로운 책들이 추가된다. 이용자들의 입장에서 이러한 사실은 다양한 장서를 접할 수 있다는 장점으로 여겨질 수 있지만 이용자들은 많아진 선택지 속에서 어떤 책을 읽을지 결정을 내리는 것에 부담을 느끼며 이러한 부담은 도서관 이용에 있어 장애물로 적용된다. 대학 도서관은 매년 상당한 비용을 지불하여 수서를 진행하며 이용자들이 이를 활발하게 활용하기 위해서는 도서 추천시스템의 필요성이 대두되고 있다.

최근 개인 맞춤형 추천시스템을 대학 도서관에 도입하려는 움직임이 활발하다. 2019년 서

울대학교 중앙도서관은 'S-Curation'이라는 명칭의 개인화 서비스를 제공하기 시작했다. 해당 시스템은 학생의 도서 대출 이력에 대한 '좋아요/싫어요' 기능과 본인이 직접 등록한 관심 키워드를 바탕으로 도서 취향을 파악한다. 이를 바탕으로 개인별로 맞춤형 도서 및 보고서를 추천해주는 방식으로 운영된다(SNU NOW, 2019). 또한, 이화여자대학교에서는 2017년부터 제공해 온 '독서 프로파일링 서비스'를 2020년 개편하였다. 추천 도서의 범위를 전자책까지 확대하고, 추천 도서가 특정 분야로 편중되지 않도록 사용자가 관심 분야를 관리하는 기능도 추가했다(허정윤, 2021).

하지만 위에 언급한 대학교와 달리 개인화된 도서 추천 서비스가 도입되지 않은 대학 도서관들이 존재한다. 이러한 도서관의 개인화 도서 추천을 어렵게 만드는 이유 중 하나가 도서에 대한 이용자들의 평가 체계가 부재하다는 점에 있다. 그렇기 때문에 많은 도서관들의 도서 추천시스템은 이용자가 검색한 도서와 비슷한 주제에 한하여 동일 주제 내 인기 자료를 추천하거나 별도로 시행 중인 북 큐레이션과 같은 프로젝트를 통하여 사서가 직접 인기 도서를 선정하고 전시하는 제한된 방식으로 서비스가 제공되고 있다. 도서관은 이미 필수적 업무로 많은 양의 대출 데이터를 수집 및 보관하고 있다. 이러한 데이터들을 활용하여 도서에 대한 개인의 선호도를 계산할 지표를 구축하고, 도서를 추천해줄 수 있는 시스템이 만들어진다면, 도서관은 보다 효율적으로 자원을 관리하고, 많은 이용자들이 다양한 자료를 손쉽게 이용할 수 있을 것이다.

본 연구에서는 도서 평가 체계가 없는 대학

도서관에 도입이 가능한 도서 선호도 산출과 이에 기반한 개인화 도서 추천시스템 개발을 목표로 하였다. 대출 기록을 이용하여 이용자의 성향을 파악하며 이를 선호도 평가지표를 설정하고, 이용자가 아직 읽지 않은 책에 대한 예상 선호도를 산출하는 방식으로 도서를 추천한다. 이러한 추천 결과에 따른 사용자 평가를 통하여 추천시스템 모델들을 더욱 발전시킬 수 있을 것이라 생각한다.

2. 이론적 배경

2.1 추천시스템

본 연구는 대학 도서관에 적용 가능한 개인별 도서 추천시스템을 구현하는 것을 목표로 한다. 추천시스템의 구현방식은 크게 협업 필터링(Collaborative Filtering), 행렬 분해(Matrix Factorization) 그리고 딥러닝(Deep Learning)을 사용한 방식으로 분류할 수 있다. 본 연구에서는 구현의 난이도가 상대적으로 복잡하지 않고 실제 대학 도서관 도서 추천시스템 개발 및 사용까지 IT 기반시설 등에 대한 초기비용이 많이 들지 않는 방안인 협업 필터링 기법 중 사용자 기반 협업 필터링 기법과 행렬 분해 기법을 사용하였다.

2.1.1 사용자 기반 협업 필터링(User Based Collaborative Filtering)

사용자 기반 협업 필터링 알고리즘은 사용자의 아이템에 대한 선호도를 기반으로 유사한 사용자들의 그룹을 찾아내고, 사용자들 사이의

아이템 선택 패턴의 유사성을 통해 아이템의 선호도를 예측한다. 이는 사용자의 아이템 선호도가 유행을 따르거나, 일정한 패턴을 가지고 있다는 사실에 근거하고 있다(Herlocker et al., 2004). 유사한 사용자 그룹에서 선호도가 높은 아이템들을 추천해주는 방식이 사용자 기반 협업 필터링 기법을 사용한 추천 방식이다. 사용자 기반 협업 필터링에서 가장 중요한 부분은 사용자 사이의 유사도를 측정하는 것이다. 사용자가 아이템에 대하여 남긴 선호도 기록을 활용하여 사용자와 사용자 간의 유사도를 계산하고 이를 기반으로 알고리즘을 구현한다. 유사도 측정에는 다양한 방법들이 있다. 피어슨 상관관계 기반 유사도 측정, 유클리드 거리 기반 유사도 측정, 코사인 유사도 측정 등이 있다. 유사도 측정 방법에 따라 추천시스템이 성능이 달라질 수 있으므로, 추천 환경에 맞는 유사도 방법을 적절하게 사용하여야 한다(조연선, 2019).

2.1.2 행렬 분해(Matrix Factorization)

행렬 분해란 사용자-아이템 행렬(User-Item Matrix)을 사용자 잠재요인 행렬(User-Latent Matrix)과 아이템 잠재요인 행렬(Item-Latent Matrix)의 곱으로 분해하는 행렬 인수분해 방법이다. 사용자가 평가하지 않은 아이템에 대한 선호도를 쉽게 추정할 수 있는 방법으로 Singular Value Decomposition(SVD), Stochastic Gradient Descent(SGD), Alternating Least Squares(ALS), 마르코프 체인과 같은 알고리즘들이 주로 사용된다(Hu, Koren, & Volinsky, 2008).

SVD는 특이값 분해 방법이다. 이는 행렬의 전체적인 구조를 기반으로 추천을 하는 방식이다. 차원 축소 개념의 일종으로 고차원 행렬을

저차원의 행렬로 축소 시켜 분석의 정확성을 높이고 계산 속도를 향상 시킬 수 있다. SVD는 기존의 행렬을 직각 행렬 2개와 1개의 대각 행렬로 분해한다. 추천시스템에서는 각각의 분해된 행렬을 이용하여 차원 축소를 실시할 수 있다. SVD는 평가 정보가 있을 경우 이를 바탕으로 행렬을 분해하였을 때, 잠재 요인을 잘 정의한다고 알려져 있다(정승운, 2017).

SGD 방법은 실제 선호도와 예측된 선호도의 차이를 에러로 정의하며 이는 결측치와 상관없이 선호도 계산을 진행한다는 장점이 있다. 우선, 값이 존재하는 선호도를 이용해 순차적으로 경사 하강법(Gradient Descent)을 진행하며 오차를 계산한다. 이 과정에서 편미분을 통해 사용자 잠재요인과 아이템 잠재요인의 업데이트를 진행해 선호도를 산출한다(Koren, Bell, & Volinsky, 2009).

2.2 도서 추천시스템

대학 도서관의 효율적인 장서 활용을 위한 도서 추천과 사용자의 개인적 특징 및 취향들을 접목한 도서 추천시스템에 관한 다양한 연구가 이뤄지고 있다. 개인의 대출 기록을 반영한 도서 추천시스템의 필요성에 대해 연구한 논문들은 다음과 같다.

2.2.1 추천 도서의 기준 설정 관련 연구

박양하(2016)는 추천 도서 목록의 효과적인 사용을 위한 정보서비스로 학교도서관의 북 큐레이션을 제안하고 큐레이션의 기준이 되는 도서의 추천 기준을 도출하고자 하였다. 이를 위하여 추천 도서 목록을 분석하여 이용자 정보

와 도서 정보에서 추천에 활용할 수 있는 속성에 대하여 분석하였다. 이러한 분석의 결과로 큐레이션 기준에 대한 중학생의 상위 3개 선호 항목으로 '관심키워드 중심 추천', '학년단위 다대출', '교과연계도서목록 다대출' 고등학생의 상위 3개 선호 항목으로 '관심키워드 중심 추천'과 '학년단위 다대출', '교과연계도서목록 다대출'이라는 추천 기준을 도출하였다.

Tsujii, Takizawa et al.(2014)은 도서 추천의 기준으로 도서관 대출 기록과 도서 콘텐츠에 대한 정보를 서포트 벡터 머신을 활용하여 이용자에게 추천 도서 목록을 제공하고 이에 대한 평가를 진행하였다. 세부적으로 해당 연구에 활용된 데이터로는 대출 기록에 근거한 신뢰와 연관 규칙에 대한 지원, 도서 제목 간의 유사성, 도서 일본 소수 분류(NDC) 범주 간의 일치/불일치, BOOK 데이터베이스에 있는 도서 개요 간의 유사도 등을 활용하였다.

2.2.2 도서 추천 모델 구현 관련 연구

노영희(2014)는 클라우드 서비스, 빅데이터, 상황인식 기술, SNS를 활용한 도서관 주제 봉사서비스, 이용자 맞춤형 서비스, 도서 추천 서비스, 맞춤형 이용자 교육 서비스 등 미래 도서관에 사용될 수 있는 기술들을 검토하고 차세대 디지털 도서관에 활용될 수 있는 개념과 적용 사례들에 대한 분석을 진행하였다.

정희정, 조성배(2011)는 도서관 정보시스템에서 보다 개인화된 추천 서비스를 제공하기 위하여 사용자 기반 협업 필터링을 활용한 방안을 제시한다. 아이템에 해당하는 도서를 주제분류번호를 이용하여 동일 주제의 자료끼리 군집화하고 주제 군집에 대한 선호도 점수를

추출하여 이를 사용자 유사도 계산에 사용하는 방식을 도출했다.

Tsuji, Yoshikane et al.(2014)은 도서관 대출 기록을 포함한 몇 가지 기능을 기반으로 머신러닝 모듈을 통해 책을 추천하는 방법을 제안한다. 서포트 벡터 머신과, 랜덤 포레스트, Adaboost 등의 머신러닝 모듈을 활용하여 도서 추천 서비스를 설계하였다.

Kanetkar et al.(2014)은 학생들의 도서 선택을 하나의 과제로 보고, 일반적인 협업 및 콘텐츠 기반 필터링 접근법과 별도로 권장사항을 제공하는 다양한 측면을 활용하는 웹 기반 개인화된 하이브리드 북 추천자 시스템을 위한 모델을 제시하였다.

위에서 본 바와 같이 도서관의 효율적인 자원 사용을 위한 도서 추천 서비스 관련한 다양한 분야의 연구들이 진행되었다. 하지만 도서관의 사용자 및 도서 데이터를 사용하여 추천 목록을 생성하고 사용자 기반의 평가를 진행한 실질적인 도서 추천시스템을 제안한 연구는 국내에서 진행되지 않았다. 따라서 본 연구는 다양한 추천 기법들을 통하여 이용자의 개인적 취향을 고려한 도서 추천 목록을 도출한 후 이를 실제 이용자에게 평가받았다.

3. 연구방법

3.1 연구절차

본 연구에서는 대학 도서관 개인화 추천시스템 제안을 목적으로 성균관대학교 학술정보관의 데이터를 활용하여 분석 및 시스템 설계를 진행

하였다. 추천시스템 설계에 필요한 선호도는 도서 장르에 의한 선호도와 대출 기간에 의한 선호도를 활용하여 계산하는 방식을 개발하였다.

3.2 연구설계

3.2.1 데이터 전처리

본 연구에서는 성균관대학교 학술정보관의 2019년도의 학부생 대출 기록 데이터를 대상으로 분석과 시스템 구축을 진행하였다. 대출 기록은 대출자의 정보, 대출 도서의 정보, 대출이 이루어진 시점과 장소 등에 대한 정보로 이루어져 있다. 그 중 대출자 번호, 대출자 소속, 도서명, 청구기호, 대출일자, 반납일자만을 예측에 사용하였다. 자격증 시험 대비 문제집과 같은 취업에 도움이 되는 서적들과 과제도서로 분류된 도서들이 개인의 기호를 반영한다고 보기에는 무리가 있어 해당 도서들은 제외하였다. 또한, 대출자의 도서 선호도 예측을 위해서는 충분한 도서 대출 기록이 있어야 하므로 도서를 한 권 대출한 학생과 소속 학과가 없는 학생들의 기록은 삭제하였다.

대출자 번호의 경우 개인정보의 유출을 방지하기 위해 난수 처리된 상태로 제공 받았으며 단과 대학, 특정 학과 등의 다양한 기준으로 분류되어 있던 대출자 소속은 단과대 단위(문과 대학, 사회과학대학, 경영대학 등)로 재분류하였다.

도서관 내의 도서는 이용자들이 동시에 동일한 도서를 대출할 수 있도록 복본을 비치해 둔다. 복본은 서지정보, 판 사항, 제목 등 모두 동일하지만 청구 기호를 통해 복본인 것을 나타낸다. 모든 도서는 고유한 등록번호를 갖고 있

어 본 연구에서 등록번호를 통해 도서를 구별 하는데, 복본을 동일한 도서로 인식할 수 있도록 기준이 되는 도서와 복본의 등록번호를 동일하게 연결 시켜주었다.

해당 대학 도서관으로부터 제공된 데이터 변수명과 내용은 <표 1>과 같다.

<표 1> 제공 데이터 상세

변수명	내용
등록번호	학술정보관 도서 등록번호
대출처리일자	대출이 처리된 날짜
반납처리일자	반납이 처리된 날짜
학번 난수	난수 처리한 학번
소장분관	도서가 소장된 분관
소장서고	도서가 소장된 서고
대출상태	대출중 / 반납
대출유형	장기대출 / 대출취소
반납유형	정상반납 / 연체반납
서명	도서 제목
청구기호	도서 청구기호(DDC)
서지번호	도서 정보를 식별하는 번호
전공	대출자가 소속된 학과

3.2.2 도서 선호도 계산

해당 대학 도서관에는 도서에 대한 이용자의 선호도를 파악할 수 있는 지표가 없으므로 선호도를 계산하기 위한 식을 직접 도출하였다. 본 연구에서는 도서 장르, 대출 기간을 이용하여 선호도를 계산하는 방법을 제안한다. 이 두 가지 방법을 각각 사용하여 모델을 구축하거나 여러 개의 기준을 조합하여 선호도를 계산하는 방법 등 다양한 방법으로 활용될 수 있다.

3.2.2.1 도서 장르에 의한 선호도

도서 장르에 의한 선호도란 특정 장르를 선호하는 이용자들의 선호도를 추천시스템에 반영하

기 위한 지표이다. 해당 이용자가 대출한 전체 도서의 장르 중에서 해당 도서의 장르의 비율을 파악하여 이를 5개의 구간으로 나누어 선호도 점수를 매기는 방법을 제안한다. 이때 특정 장르의 비율은 백분율로 나타내며 0%~20%는 1점, 20%~40%는 2점, 40%~60%는 3점, 60%~80%는 4점, 그리고 80%~100%는 5점으로 총 5개의 구간으로 나누어 선호도를 부여하였다.

대학 도서관의 경우 듀이 십진 분류표(DDC)를 도서 분류 체계로 이용하기 때문에 DDC를 기준으로 이용자의 장르 선호도를 파악할 것을 제안한다. DDC 기준으로 10의 자리 숫자까지 같은 책들을 같은 장르의 책으로 판단하기로 했다. 예를 들어 8로 시작하는 책은 문학으로 분류된 책이다. 두 번째 자리가 2인 경우 영미문학으로 82X인 경우 영미 문학에 속하는 도서이다. 같은 문학책이더라도 세부적인 취향 차이가 있을 수 있기 때문에 10의 자리 수까지 같은 책을 동일한 장르로 판단했다.

DDC의 예시는 <표 2>와 같다.

<표 2> DDC 분류

800	문학
810	동양문학
820	영미문학
821	시
822	희곡
823	소설
830	독일문학
840	프랑스문학
850	이탈리아문학
860	스페인문학
870	라틴문학
880	회람문학
890	기타문학

3.2.2.2 대출 기간에 의한 선호도

대출 기간에 의한 선호도란 도서를 대출한 기간이 길수록 도서에 대한 선호도가 높다고 평가하는 지표이다. 오승선(2015)은 대출 기간이 0일이면 1점, 평균일보다 적으면 2점, 대출 기간이 평균일과 같으면 3점, 대출 기간이 평균일+7일 이내이면 4점, 대출 기간이 평균일+7일 이상이면 5점을 부과하였다. 본 연구는 개인의 도서 이용 행태에 따라 대출 기간이 달라지므로 단순히 대출 기간으로 선호도를 도출하기는 어렵다고 판단하여 개인의 평균 대출 기간을 이용한 식을 도출했다.

$$\text{대출 기간 선호도} = \frac{\text{특정 도서 대출 일수}}{\text{해당 대출자의 평균 대출 일수}}$$

우선, 특정 대출자가 평균적으로 책을 읽는 평균 대출 일수를 구한다. 특정 도서를 대출자가 며칠 동안 읽었는지를 구하여 평균 대출일수로 나눠준 것을 대출 기간 선호도로 나눠주었다. 만약 특정 도서를 대출자의 평균 대출 일수만큼 대출을 하였다면 선호도는 3, 평균 이하로 하였다면 구간에 따라 1~2, 평균 이상으로 하였다면 4~5점을 부여한다. 특정 도서의 대출 일수가 평균 이하일 경우, 특정 도서 대출 일수를 평균 도서 대출 일수로 나눈 결과값의 구간이 0이상 0.5 미만 일 때 1점, 0.5 이상 1 미만 일 때 2점으로 산출한다. 또한 특정 도서의 대출 일수가 평균 이상일 경우, 결과값의 구간이 1.0 초과 1.5 이하 일 때, 4점 1.5 초과 2.0 이하 일 때, 5점으로 각각의 선호도를 산출하였다.

3.2.2.3 최종 도서 선호도

최종 도서 선호도는 도서 장르에 의한 선호도와 대출 기간에 의한 선호도를 모두 활용하여 최종 지표를 산출하였다. 해당 연구는 도서 장르에 의한 선호도를 50%, 대출 기간에 의한 선호도를 50% 반영하여 최종 선호도를 도출하였다. 해당 비율은 상대적인 것으로 대학 도서관별 특성에 따라 비율을 조절하여 활용할 수 있다.

$$\text{최종 도서 선호도} = \text{장르 선호도} * 0.5 + \text{대출 기간 선호도} * 0.5$$

대출자를 대표하는 난수, 해당 도서의 등록번호, 그리고 해당 도서에 대한 최종 선호도를 추천시스템에 사용할 1차적인 데이터 형태로 사용하였고, 예시는 <표 3>과 같다.

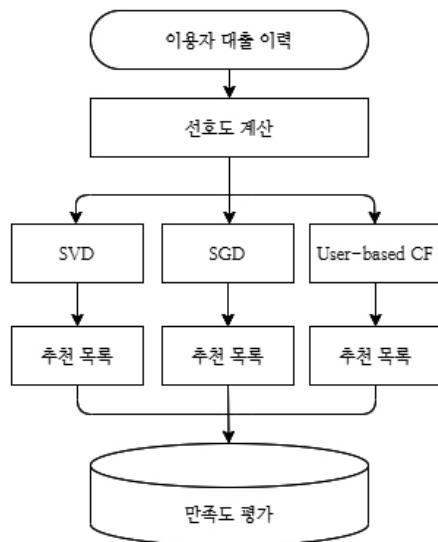
<표 3> 최종 선호도 데이터 프레임 예시

학번 난수	등록번호	선호도
572900	EE0462045	3.5
572900	EM0871360	4.0
991018	EE0306724	3.5
9999115980	EM0985200	2.0
9999486850	EE0360477	4.5
9999486850	WM0186535	3.0

3.3 도서 추천시스템 모델

구축한 도서 추천시스템의 입력 값은 이용자 개인의 도서 대출 이력이다. 이용자들의 대출 이력을 이용하여 장르, 대출기간 등을 활용한 도서에 대한 선호도를 산출한다. 이를 추천 모델에 사용하였으며 추천시스템 모델은 크게 3가지 모델로 구축하였다. SVD, SGD, User-based Collaborative Filtering 세 개의 추천 모델 모두

선호도가 필요한 모델들로 위의 최종 선호도를 모델의 입력 값으로 사용하였다. 대학 도서관에 따라서 구현하기 적합한 모델을 선택하고 추천 모델을 통하여 도서 추천 목록을 도출하게 된다. 이용자에게 추천 목록을 제공하면 사용자 평가를 받고 이 결과를 데이터베이스에 저장한다. 이를 그림으로 나타낸 것이 <그림 1>이다.



<그림 1> 도서 추천시스템 설계

4. 연구 결과

4.1 데이터셋 구축

전처리한 데이터를 이용하여 추천시스템에 사용될 데이터셋 형태로 재구축하였다. 추천시스템에 사용한 데이터는 2019년 1년간의 대출 기록으로 102,528건을 사용했다. 학부생은 총 9,168명, 도서는 총 45,175권으로 이루어져 있다. 대출 기록에 있는 도서별로 3.22의 도서 선

호도 계산 방식을 활용하여 입력 데이터를 구성했다. Python의 pivot_table 함수를 사용하여 index에는 모든 학부생들의 학번 난수, column은 모든 도서 등록 번호로 구성하였으며 value는 각 학부생의 도서에 대한 선호도를 값으로 추천시스템에 활용한 새로운 데이터셋을 구축했다. 대출하지 않은 도서에 대해서는 선호도가 0으로 채워지는 형태이다.

4.2 추천시스템 구축 및 평가

4.2.1 행렬 분해 기반 추천시스템

2.1.2절에서 설명한 SVD와 SGD 기법을 사용해 모델링을 진행했다. SVD 모델의 경우 베스트셀러들이 추천 도서 목록으로 자주 도출되었다. 다른 이용자들의 선호도가 높은 도서들을 중심으로 개인의 독서 행태와 장르 선호도에 따라 추천 목록이 조금씩 변형되는 것을 확인했다. User-Item Matrix가 결측값이 많은 희소 행렬이기 때문에 다른 이용자들의 영향을 크게 받는 것으로 나타났다. 행렬 분해 기반 모델 평가 지표로 Root Mean Square Error(RMSE)를 이용하였다. RMSE는 모델의 예측값과 실제 환경에서 관측된 값의 차이를 계산하는 척도이며 값이 작을수록 정밀도가 높다고 볼 수 있다. SVD 모델의 RMSE는 0.75가 도출되었다.

SGD 알고리즘을 사용한 결과 RMSE는 0.0032이 도출되었다. SVD와 다르게 오차를 계속해서 업데이트하여 반영하기 때문에 베스트셀러 위주로 추천 목록이 도출되는 문제점은 해결할 수 있었다. 대출자의 장르 선호도를 SVD 모델보다 더 반영하여 개인마다 추천되는 목록이 크게 달랐다.

두 개의 모델 모두 이용자별로 예측된 결과 값이 큰 도서 중 이용자가 읽지 않은 도서 5권을 추천하는 방식으로 진행했다.

용자와 유사도가 높은 학생부터 5권의 책을 선정하여 추천해주었다.

4.2.2 User-based Collaborative Filtering

기반 추천시스템

최종 선호도 설정 후, 이용자와 이용자 사이의 유사도 계산을 위하여 학생들의 책에 대한 선호도 행렬을 만들었으며, 행에는 학부생 학번 난수, 열에는 학부생들의 대출 기록에 있는 도서들의 등록번호가 포함되어 있다.

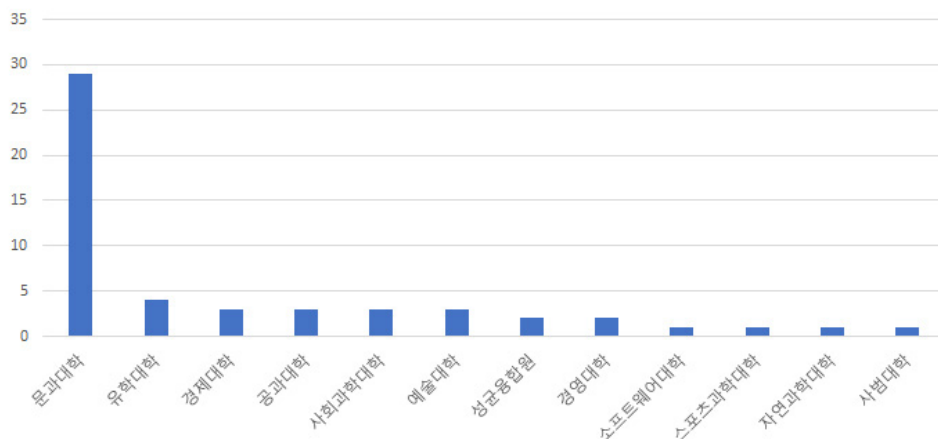
최종 선호도 행렬에 대하여 코사인 유사도 측정 방식을 사용하여 학생들 사이의 유사도 행렬을 만들었다. 만들어진 유사도 행렬에 기반하여 특정 학부생의 학번 난수 번호를 입력하면 이와 유사도가 높은 학부생 10명이 포함된 그룹을 추출하였다. 유사도가 높은 학생들 그룹이 읽었던 책들을 10개씩 추출하였다. 추출된 책들 중 추천받을 이용자의 대출 기록에 있는 책들을 제거하였고, 추출된 책들의 중복을 삭제하여 목록을 정리하였다. 사용자 평가를 위하여 목록 중 이

4.3 사용자 평가

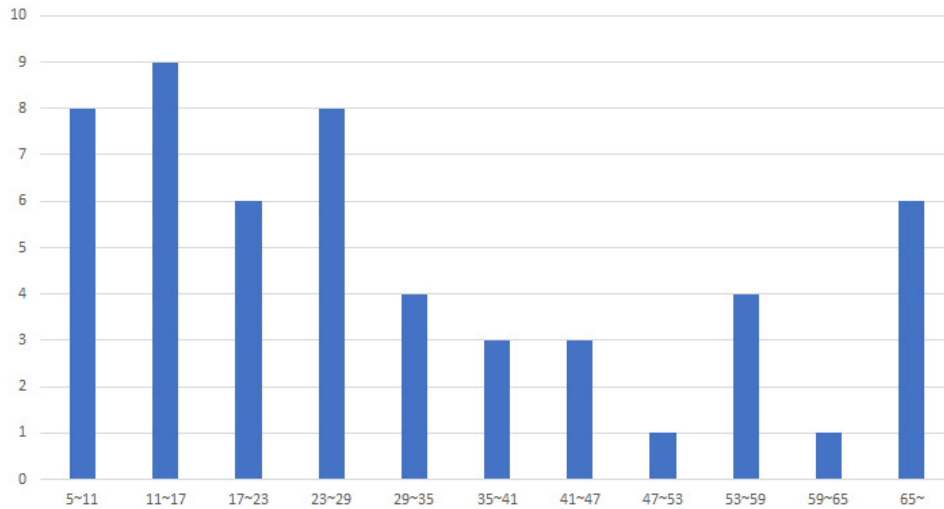
4.3.1 실험 대상자의 특성

SVD, SGD, 사용자 기반 협업 필터링 세 가지 모델의 평가를 53명의 다양한 소속의 학부생들을 실험 대상으로 선정하여 사용자 평가를 진행했다. 추천시스템 모델들에 활용된 선호도 지표는 앞서 3.2.2.3절에서 언급한 최종 선호도 지표를 활용하였다. 아래 실험 대상자 단과 대학 분포는 대상자들의 대학이 많이 분포한 순으로 그래프로 보여준 결과이다. 문과대학은 29명, 유학대학 4명, 경제대학 3명, 공과대학 3명, 사회과학대학 3명, 예술대학 3명, 성균궁합원 2명, 경영대학 2명, 소프트웨어대학 1명, 스포츠과학대학 1명, 자연과학대학 1명, 사범대학 1명 순으로 실험 대상자들의 대학이 분포했다(〈그림 2〉 참조).

〈그림 3〉은 실험 대상자들의 도서 대출 권수



〈그림 2〉 실험 대상자 대학 분포



〈그림 3〉 실험 대상자 대출 권수 분포

분포를 살펴본 차트이다. 11권에서 17권을 읽은 대상자들은 9명으로 가장 보편적으로 읽은 대출 권수이다. 다음으로 많이 분포한 범위는 5권에서 11권, 23권에서 29권이다. 이어서 65권 이상을 읽은 실험 대상자와 17권에서 23권을 읽은 실험 대상자는 6명, 29권에서 35권, 53권에서 59권을 읽은 실험 대상자는 4명이 분포했다.

4.3.2 도서 추천에 대한 만족도

실험 대상자들의 대출 이력을 입력값으로 넣어 각 모델에서 도출한 예측값 중 상위 5개의 도서를 추출했다. 평가지는 각 모델에 대한 추천 목록, 도서별 등록번호, 서명, 저자, 발행사향, 형태사향, 표지를 삽입하여 도서에 대한 정보를 제공하였다. 실험 대상자들은 추천된 도서 목록 중 읽고 싶은 책에는 '만족', 그렇지 않은 책에는 '불만족'을 부여하는 방법으로 평가했다.

실험 대상자의 대출 이력에 따른 3가지 모델

에 따른 추천 목록은 〈그림 4〉와 같고, 실험 대상자의 피드백은 〈그림 5〉와 같다.

실험 대상자들의 평가를 바탕으로 각 모델의 사용자 만족도를 평균 내어 산출한 결과는 〈표 4〉와 같다. 실험 대상자들의 평가는 각 모델에 대한 5개의 도서 추천 목록에 각각 만족, 불만족 표시를 하며 이는 만족인 경우 1점, 불만족인 경우 0점을 주어 한 모델에 대해 5점 만점의 만족도 점수로 환산한다. 실험 대상자들의 추천 목록에 대한 만족도가 가장 높았던 행렬분석기법의 기반인 SVD 모델로, 5점 만점에 3.6점으로 3개의 모델 중 가장 높은 만족도를 보였다. SVD 모델에 비해 개인의 선호도를 더욱 반영할 것으로 예상했던 SGD 모델은 2.6점의 만족도로 가장 낮은 만족도를 보였다. 사용자 기반 협업 필터링의 경우 3.15점으로 SVD 모델에 비해서는 낮은 만족도, SGD 모델보다는 높은 만족도를 보였다.

정량적 평가인 만족, 불만족 평가 외에 실험

자료명	등록번호	저자	발행처
(데이터 분석을 위한) 파이썬 철저 입문 : 기초 문법부터 실무에 필요한 데이터 분석 기술까지 한 번에 배우는 파이썬을 이용한 데이터 분석 : 다양한 파이썬 라이브러리와 소프트웨어를 활용한 데이터 사이언스의 이해	EM1025957 EM1018588	최은석 Fandango, Armando	파주 : 위키북스, 2018 서울 : 에이콘, 2018
Data communications and networking 마감수, 금기와 위반의 상상력 허버드 대학의 시계열교육 : 4~8개월간의 그래픽디자인 연습 중형무진 역사 : 한국사 동양사 서양사를 함께 읽는다 몸, 한의학으로 다시 태어나다 : 한의학으로 밝힌 우리 몸 건강백과 R과 통계분석 R을 이용한 빅데이터분석 입문 (교양으로 읽는) 우리 몸 사전 몸, 한의학으로 다시 태어나다 : 한의학으로 밝힌 우리 몸 건강백과 (처음 만나는) 파이썬 선형대수학 =Linear Algebra 빅뱅이론과 우주의 기원 =The big bang theory & the origin of the universe (Do it! 쉽게 배우는) R 데이터 분석 =Do it! R for data analysis 이토록 철학적인 순간 : 내 인생의 20가지 통과의례 : 자전거타기에서 컷스까지, 학교에서 이사까지 (Hackers) TOFL actual Test : Reading 조선명저기행	WM0250682 EM1045925 EM0541539 EM0938011 EM0910043 EE0502546 EM1012041 EM1013974 EM0910043 EM1003049 EM1000839 EM0967799 EM1010197 EM0953265 CL0006298 EM1019631	Forouzan, Behrouz A 고운기 가따야마, 도시히로 남경태 안세영 박동린 최민석 안세영 가따야마, 마사히로 강병련 윤성민 김영우 Smith, Robert Rowland 해커스 어학연구소 박영규	New York : McGraw-Hill, c.2013. 서울 : 역락, 2019 서울 : 국책, 1993. 서울 : Humanist, 2014 서울 : 와이즈리, 2010 파주 : 자유아카데미, 2019 서울 : 한티미디어, 2017 파주 : 서해문집, 2017 서울 : 와이즈리, 2010 파주 : 제이컴, 2017 대전 : 공미디어 : 충남대학교출판문화원, 2017 서울 : 부지런한 거북이 : 대광서림, 2015 파주 : 이지스퍼블리싱, 2017 서울 : 웅진지식하우스 : 웅진씽크빅, 2014 파주 : 해커스 어학연구소, 2016 파주 : 김영사, 2018



User-based CF				SVD				SGD					
	등록번호 E3030054	제목 300년간의 역사 탐험정기 1:가죽과 함께 일출 동굴 지난 역사에서 통찰여행기	저자 정병	발행사 서울 : 뉴비, 2004.	등록번호 E30302359	제목 A만아 공자 인간 생활소설	저자 한강, 1970-	발행사 최은영, 2014	등록번호 E30079223	제목 [백은 공자를 두 배로 높여준노] 정지혜 소설 이야기 343	저자 정지혜	발행사 서울 : 에디스, 2016	
	등록번호 E300912053	제목 (41권과 42권으로 분산) 반한 반한	저자 서용석	발행사 서울 : 웅진지식하우스 : 웅진씽크빅, 2013		등록번호 E30096376	제목 세익스피어와 인간 생활소설	저자 한강, 1970-	발행사 최은영, 2007		등록번호 E300612054	제목 한국현대문학사	저자 김영민
	등록번호 E30338487	제목 [Forouzan]의2권(데이터통신)	저자 Forouzan, Behrouz A.: 차재용	발행사 서울 : 한국 계조출 : McGraw-Hill Korea, 2006.		등록번호 E30100518	제목 소직 두 사람 [인생의 소설]	저자 김영하	발행사 최은영, 2017		등록번호 E301020048	제목 David S. Moore, William I. Note	저자 David S. Moore, William I. Note
	등록번호 E30094039	제목 이산수학	저자 Johnsonbaugh, Richard, 김호일, 채재혁 : 이경환 등	발행사 (서울) 에이치앤에듀케이션코리아 : 한티미디어, 2008		등록번호 E30102019	제목 살인의 문	저자 최지나노, 정지호 : 이계태	발행사 서울 : 예인, 2018		등록번호 E300712723	제목 사망할 준비를 하라니	저자 윤희태
	등록번호 E30090902	제목 [Rozen]의 사탕수박	저자 Rozen, Kenneth H.: 공은혜, 왕영진 : 안영환 등	발행사 서울 : 한국 계조출 : McGraw-Hill Korea, 2017		등록번호 E3001286	제목 연차의 모든 것	저자 최지나노, 정지호, 1956- : 김단우, 1958-	발행사 서울 : 예인, 2019		등록번호 E301006642	제목 두 발로 걷 : 인체의 양면도	저자 김희진

<그림 4> 실험 대상자의 평가지

추천 시스템에 대한 만족도와 전체적인 피드백 부탁드립니다(자유롭게 서술).

추천 받은 책에 대한 정보가 부족해 내가 이 책이 정말 필요한 책인지에 대해서 확신할 수 없을 때가 있다. 서지사항 뿐만 아니라 내용에 대한 심도 깊은 추천이 더해졌으면 좋겠다.
그렇지만 전반적으로 관심 분야에 대한 파악이 잘 된것 같다. 몇몇 작품들은 기호에 맞지 않았으나 그 외의 대부분은 내가 평소에도 흥미를 가지고 독서를 하던 분야의 것들이었다

<그림 5> 실험 대상자의 피드백

〈표 4〉 사용자 평가 결과

	User-based CF	SVD	SGD
평균	3.15	3.6	2.6
중간값	3	4	2
표준편차	1.22	1.43	1.38

대상자들을 대상으로 해당 추천시스템에 대한 의견을 묻는 정성적 평가도 진행하였다. 실험 대상자들의 대학 도서관 대출 기록만으로 도서 추천 목록을 제공하였지만, 개인적으로 외부에서 구매하여 읽은 도서가 추천 목록에 포함되었다는 점에서 높은 만족도를 보였다. 또한, 추천 목록에 대해서 높은 신뢰도를 보였다. 이와 같이 긍정적인 의견이 대부분이었지만 한계점 및 개선될만한 사항 또한 알 수 있었다. 도서의

출판사나 표지 등 더욱 다양한 고려 요소를 반영하지 못하여 이러한 요소들을 대출 시 중요하게 판단하는 대출자들의 선호를 고려하지 못했다는 점에서 아쉽다는 평가가 있었다. 또한, 추천된 이유도 함께 제공하는 것이 좋겠다는 개선 사항도 있었다.

의견을 제시한 실험 대상자는 전체 53명 중 19명이었고, 사용자 피드백의 일부분을 〈표 5〉로 나타내었다.

〈표 5〉 사용자 피드백

구분	피드백
긍정적인 평가	학교 도서관을 이용하다보니 내가 읽고 싶은 책만 빌리게 아니고 과제나 시험을 위해서 관심없는 분야의 책도 꽤 빌렸었다. 그래서 그동안 대여한 책들의 주제나 장르가 중구난방이었는데도 생각보다 추천도서 중 읽고 싶은 책이 많아서 만족스러웠다. 심지어 추천 도서 중 이미 가지고 있는 책이 두권이나 있길래 추천 프로그램에 대해 신뢰할만하다고 느꼈다.
	실제로 읽은 책도 있고 넘 신기해요.
	추천해주신 도서들 대부분이 마음에 들었습니다. 실제로 읽어본 책도 있고 읽고 싶다고 생각했던 도서들도 있어서 굉장히 신기했습니다. 주로 과제용으로 책을 대출해서 그런지 제 전공 관련 책들도 추천돼서 더 좋았습니다.
개선 사항	책을 읽을 때 호불호가 많이 갈리는 편이라서 누가 내 취향에 맞는 책들을 추천해주면 좋을 것 같다는 생각을 하곤 했습니다. 제가 제공한 표본이 그렇게 풍부하지도 않았는데, 생각보다 훨씬 읽어보고 싶고 관심이 가는 책들을 많이 추천해줘서 놀라웠습니다. 심지어 추천해준 책 중에서 세권은 이미 읽었고 재미있게 본 기억이 있어서 더 만족스러웠네요.
	한 출판사가 지향하는 주제, 소재가 있고 그리고 유지하고자 하는 퀄리티가 있습니다. 그 때문에 작가 만큼이나 중요한 의미를 가진다고 생각합니다. 이러한 부분이 고려되었는지는 모르겠지만 고려된다면 좋을 것 같습니다.
	전반적으로 매우 만족스럽습니다. “한번 읽어볼까?” 생각이 드는 추천도서가 많았습니다. 다만 왜 이 책이 추천됐는지 너무 궁금해서 이후에 서비스화 된다면 추천이유를 알 수 있는 서비스가 포함되었으면 좋겠어요.
	청구기호나 서고위치 정보도 같이 제공되면 좋겠습니다.

5. 결론

본 연구에서는 대학 도서관 이용자 개인의 대출 기록 데이터를 이용하여 개인화 맞춤 도서 추천시스템을 구현했다. SVD와 SGD 기법을 이용한 행렬 분해 모델, 사용자 간의 유사도 계산을 통해 도서를 추천하는 사용자 기반 협업 필터링 총 3가지 모델링을 진행했다. 객관적인 모델의 비교 및 평가를 위해 해당 대학 도서관 이용자를 대상으로 사용자 만족도를 조사했다. 결과적으로 가장 만족도가 높은 모델은 5점 만점의 3.6점으로 SVD 기법을 활용한 모델이었고, SGD 기법은 2.6점으로 가장 낮은 만족도를 보였다. 또한, 협업 필터링 기법은 3.15점의 만족도를 보였다.

일반적으로 독서의 특성상 영화와 다르게 별점을 매기는 등의 사용자 선호도 측정 과정이 존재하지 않기 때문에 그러한 점을 극복하고, 대출 기록만을 이용해 다양한 방식으로 개인화 맞춤 도서 추천시스템을 구현하려고 했다는 점에서 의의가 있다.

그러나 본 연구의 모델 구축 환경의 미비로 일년 치의 대출 기록만을 사용했다는 점에서 한계가 있어 대용량의 연산이 가능한 Parallel SGD 알고리즘인 DSGD, FPSGD 등을 활용해 볼 것

을 이후의 개선 방안으로 제시한다. 덧붙여 사용자 기반 협업 필터링 기법은 유사한 사용자를 뽑아, 유사 사용자 그룹 내의 대출 기록이 있는 책들을 추천했다는 점에서 책의 유사도를 고려하지 못한 한계가 존재한다. 이와 관련하여 내용 기반 협업 필터링 기법의 특징을 함께 사용하는 혼합 협업 필터링 기법의 사용을 제안한다. 또한, 구현한 모델들이 실제로 효용성이 있는지에 대한 사용자 검증을 진행하였으나, 표본이 충분하지 않다는 한계가 있다. 추후 다양한 소속의 학술정보관 이용자들을 대상으로 추가적인 검증을 진행함과 동시에, 평가 지표를 보완하여 모델의 효용성 평가를 개선할 수 있을 것이다.

다양한 분야와 플랫폼에서 사용되고 있는 추천시스템은 도서관계에서도 활발하게 연구 및 도입이 이루어지고 있다. 그러나, 다른 분야와 달리 이용자가 직접적으로 평가하는 도서 선호도가 없다는 점에서 어려움을 겪고 있다. 본 연구는 사용자의 추가적인 선호도 입력 없이 대출 기록만을 이용해 도서 추천시스템을 구현하였으며, 앞으로 더 효율적이고 효과적인 개인화 도서 맞춤 시스템의 발전을 위한 토대가 될 것으로 전망한다.

참 고 문 헌

- 노영희 (2014). 차세대디지털도서관의 발전방향논의에 관한 연구. 정보관리학회지, 31(2), 7-40.
<https://doi.org/10.3743/KOSIM.2014.31.2.007>
- 박양하 (2016). 학교도서관 북 큐레이션 서비스를 위한 도서추천 기준에 관한 연구. 한국도서관·정보

- 학회지, 47(1), 279-303. <http://doi.org/10.16981/kliss.47.1.201603.279>
- 오승선 (2015). SNS분석을 통한 도서관의 개인 도서 추천서비스에 관한 연구: 트위터 중심으로. 석사학위논문, 연세대학교 공학대학원 컴퓨터공학전공.
- 정승윤 (2017). Factorization machine을 이용한 추천시스템 설계. 디지털콘텐츠학회 논문지, 18(4), 707-712. <http://doi.org/10.9728/dcs.2017.18.4.707>
- 정희정, 조성배 (2011). 도서관 정보시스템을 위한 협업 필터링 기반 개인화 추천 서비스. 한국컴퓨터종합학술대회 논문집, 38(1A), 251-254.
- 조연선 (2019). 협업 필터링을 이용한 도서 추천 시스템이 학교도서관 이용에 미치는 효과 연구. 석사학위논문, 연세대학교 교육대학원 사서교육전공.
- 허정윤 (2021. 04. 05). '도서관이 살아있다'... 코로나19 상황에도 진화하는 대학 도서관. 한국대학신문, 출처: <https://news.unn.net/news/articleView.html?idxno=506906>
- Herlovker, J., Konstans, J., Terveen, L., & Riedl, J. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, 22(1), 5-53. <https://doi.org/10.1145/963770.963772>
- Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative Filtering for Implicit Feedback Datasets. 2008 Eighth IEEE International Conference on Data Mining, 263-272. <https://doi.org/10.1109/ICDM.2008.22>
- Kanetkar, S., Nayak, A., Swamy, S., & Bhatia, G. (2014). Web-based personalized hybrid book recommendation system, 2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014), 1-5. <https://doi.org/10.1109/ICAETR.2014.7012952>
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix Factorization Techniques for Recommender Systems, *Computer*, 42(8), 30-37. <https://doi.org/10.1109/MC.2009.263>
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76-80. <https://doi.org/10.1109/MIC.2003.1167344>
- SNU NOW (2019. 09. 30). 중앙도서관, 맞춤형 도서 추천 서비스 제공. 출처: <https://now.snu.ac.kr/47/2/1438>
- Tsujii, K., Takizawa, N., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Isumura, H. (2014). Book Recommendation Based on Library Loan Records and Bibliographic Information. *Procedia-social and behavioral sciences*, 147, 478-486. <https://doi.org/10.1016/j.sbspro.2014.07.142>
- Tsujii, K., Yoshikane, F., Sato, S., & Isumura, H. (2014). Book Recommendation Using Machine

Learning Methods Based on Library Loan Records and Bibliographic Information. 2014 IIAI 3rd International Conference on Advanced Applied Informatics, 76-79.
<https://doi.org/10.1109/IIAI-AAI.2014.26>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Cho, Yeon Sun (2019). A study on the effect of collaborative filtering recommendation system on the use of school library, Yonsei University, Korea.
- Chung, Hee Chung & Cho, Sung-Bae (2011). Personalized recommendation service based on collaborative filtering for library information system. Communications of the Korean Institute of Information Scientists and Engineers, 38(1A), 251-254.
- Jeong, Seung-Yoon (2017). A recommender system using factorization machine. Journal of Digital Contents Society, 18(4), 707-712. <http://doi.org/10.9728/dcs.2017.18.4.707>
- Noh, Younghee (2014). A study suggesting the development direction of the next generation digital library. Journal of Korean Society for Information Management, 31(2), 7-40.
<https://doi.org/10.3743/KOSIM.2014.31.2.007>
- Oh, Seung Sun (2015). A study on personal book recommendation service by SNS Analysis: Focus on Twitter, Yonsei University, Korea.
- Park, Yang-Ha (2016). A study on the book recommendation standards of book-curation service for school library. Journal of Korean Library and Information Science Society, 47(1), 279-303.
<http://doi.org/10.16981/kliss.47.1.201603.279>