

과학 학술정보 서비스 플랫폼에서 개인화를 적용한 콘텐츠 추천 알고리즘 최적화를 통한 추천 결과의 성능 평가

Performance Evaluation of Recommendation Results through Optimization on Content Recommendation Algorithm Applying Personalization in Scientific Information Service Platform

박성은, 황윤영, 윤정선
한국과학기술정보연구원 과학기술정보센터 정보융합연구소

Seong-Eun Park(kosen21@kisti.re.kr), Yun-Young Hwang(yyhwang@kisti.re.kr)
Jungsun Yoon(jsyoon@kisti.re.kr)

요약

본 연구는 과학 학술정보 서비스 플랫폼 이용자의 정보 검색 편의성을 확보하고 적합한 정보의 획득에 소요되는 시간을 절약하기 위하여, 운영 중인 서비스 메뉴와 각 서비스 별 콘텐츠 정보를 제공하는 알고리즘 중 콘텐츠 추천 알고리즘을 최적화하고 그 결과를 비교평가 하는 것이다. 추천 정확도를 높이기 위해 이용자의 '전공' 항목을 기존 알고리즘에 추가하였으며, 기존 알고리즘과 최적화된 알고리즘을 통한 추천 결과의 성능평가를 수행하였다. 성능평가 결과 최적화된 알고리즘을 통해 이용자에게 제공되는 콘텐츠의 적합도가 21.2% 증가함을 파악하였다. 이용자에게 적합한 콘텐츠를 시스템에서 자동 도출하여 각 서비스 메뉴 별로 제공함으로써 정보 획득 시간을 단축하고, 연구정보로서 가치 있는 연구결과물의 생명주기를 연장할 수 있는 방안이라는 데 본 연구의 의의가 있다.

■ 중심어 : | 과학 학술정보 서비스 플랫폼 | 개인화 | 추천 시스템 | 콘텐츠 추천 알고리즘 | 성능 평가 |

Abstract

In order to secure the convenience of information retrieval by users of scientific information service platforms and to reduce the time required to acquire the proper information, this study proposes an optimized content recommendation algorithm among the algorithms that currently provide service menus and content information for each service, and conducts comparative evaluation on the results. To enhance the recommendation accuracy, users' major items were added to the original algorithm, and performance evaluations on the recommendation results from the original and optimized algorithms were performed. As a result of this evaluation, we found that the relevance of the content provided to the users through the optimized algorithm was increased by 21.2%. This study proposes a method to shorten the information acquisition time and extend the life cycle of the results as valuable information by automatically computing and providing content suitable for users in the system for each service menu.

■ keyword : | Scientific Information Service Platform | Personalization | Recommendation System | Content Recommendation Algorithm | Performance Evaluation |

* 본 연구는 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 과학기술국제화사업 연구
임(2012K1A3A7A05033175)

접수일자 : 2017년 09월 29일

수정일자 : 2017년 10월 14일

심사완료일 : 2017년 10월 16일

교신저자 : 황윤영, e-mail : yyhwang@kisti.re.kr

I. 서론

정보가 기하급수적으로 생산·재생산되고 확산됨에 따라 이용자들은 다양한 경로로 많은 정보를 획득할 수 있게 되었다. 하지만 과도하게 많은 정보에 노출되면서 자신에게 적합한 정보를 찾는 데 어려움을 겪고 있다. 뿐만 아니라 자신이 원하지 않는 콘텐츠를 받아봄으로써 많은 시간을 소비하게 된다. 특히 국내의 과학 학술 정보 서비스 사이트들의 경우, 다양한 연구 분야, 자료 형태의 연구 결과를 등록일자 기준으로 제공하여 단편적인 서비스를 제공하는 한계를 가지고 있다. 이러한 기술적 한계로 인해, 이용자 입장에서는 수많은 정보 중 자신에게 특화된 정보를 확보하기 위해 필터링 등을 통해 정보에 접근해야 하며, 정보 확보에 많은 시간과 노력을 투입하게 된다. 또한 개인의 특정 연구 분야에서 개인에게 필요할 수도 있는 과거의 정보가 상위에 노출되지 않아 정보가 사장되기도 한다. 이러한 문제를 해결하기 위해 적합한 정보를 적절한 이용자에게 제공하는 추천 시스템에 대한 연구가 진행되어왔다.

본 연구의 대상이 되는 과학 학술정보 서비스 플랫폼인 KOSEN(www.kosen21.org)은 2017년 3월부터 빅데이터 처리 기술을 적용하여, 연구자 개인의 특성을 기반으로 연구자에게 선호도가 높은 서비스 메뉴와 각 서비스 별 콘텐츠 정보를 자동으로 제공하는 시스템을 개발하여 운영하고 있다. 본 연구에서는 기존의 알고리즘을 최적화하여 이용자의 추천 정확도를 높이기 위해 이용자의 '전공' 항목을 추가하였으며, 기존 알고리즘과 최적화된 알고리즘을 이용한 추천 결과의 성능평가를 수행하였다. 성능평가를 위해 현재 운영 중인 알고리즘에서 도출된 결과 값과 이용자의 '전공' 항목을 추가한 알고리즘을 이용해 도출된 결과 값을 이용자에게 직접 제시하여 본인과의 관련성을 평가하도록 하였다.

이러한 알고리즘 최적화에 대한 연구를 통해 서비스 이용자의 정보 검색 편의성을 확보하고, 적합한 정보의 획득에 소요되는 시간을 절약할 수 있다. 또한 개인의 특정 연구 분야에서 개인에게 필요할 수도 있는 과거의 정보가 상위에 노출되지 않는 문제점을 해결하여, 연구 정보로서 가치 있는 연구결과물의 생명주기를 연장할

수 있다는 점에서 의의가 있다. 본 논문의 구성은 2장에서 추천 시스템의 정의와 종류 및 특성 등의 선행 연구를 소개하고, 3장에서는 기존에 운영 중인 추천 시스템과 최적화된 추천 시스템을 설명한다. 그리고 4장에서는 최적화된 알고리즘의 성능평가 방법과 그 결과를 서술한다. 마지막으로 결론에서는 본 연구의 의의 및 향후 연구방향에 대해 기술한다.

II. 이론적 배경

1. 추천 시스템

1.1 추천 시스템의 정의

추천 시스템(Recommendation system)은 이용자가 콘텐츠를 이용할 때 수많은 콘텐츠 중 이용자 혹은 콘텐츠가 가지고 있는 정보를 이용하여 이용자에게 적합한 콘텐츠를 선별하여 제시해주는 것으로서, 수많은 정보 속에서 이용자들이 직접 원하는 콘텐츠를 찾아내는 수고를 줄여주는 것이다[1]. 추천 시스템에 관한 연구는 1990년대 후반 뉴스 및 웹페이지 등을 대상으로 개인화 서비스를 위한 필터링 기법으로부터 연구 개발이 시작되어, 2006년 미국 Netflix 사에서 주최한 경연대회를 기점으로 추천 기법들이 정교화, 고도화 되었다[2]. 이후에도 추천 정확도를 향상시키고 서비스 효과성을 높이기 위한 연구가 지속적으로 진행되고 있다.

이용자에게 적절한 추천 서비스를 제공하기 위해서는 이용자가 가지고 있는 특성을 파악하고 해당 이용자의 특성과 매치되는 특성을 가진 콘텐츠를 선별해내야 한다. 이를 위해 군집화(Clustering) 기법이 쓰이고 있으며, 군집화 기법은 이용자와 이용자, 콘텐츠와 콘텐츠 개체들이 가지고 있는 특성 값을 수치적으로 계산하는 유사도 계산을 통해 개체들 간에 가장 차이가 작은 개체들을 묶는다[1].

1.2 추천 시스템의 종류 및 특성

추천 기법은 비교적 오래 전부터 연구되어 왔는데, 대표적인 방법들이 인구통계학적 추천(Demographic recommendation), 콘텐츠 기반 필터링(Content based

filtering) 기법, 협업 필터링(Collaborative filtering) 기법, 또한 각각의 방법이 가지고 있는 한계점을 보완하기 위해 두 방법의 장점을 취합하는 하이브리드 필터링(Hybrid filtering) 기법이 있다[3]. 콘텐츠 추천 시스템에 대표적으로 사용하는 알고리즘은 콘텐츠 기반 추천 알고리즘과 협업 필터링 알고리즘이다. Netflix 경연대회의 결과 협업 필터링의 접근 방식을 택한 기법들이 정확도 측면에서 콘텐츠 기반 필터링 기법들보다 월등한 결과를 보였다는 점으로 인하여, 최근 연구는 협업 필터링에 대한 연구가 대다수를 차지한다[2].

1.2.1 인구통계학적 추천 기법

인구통계학적 추천 기법은 가장 기초적인 형태의 추천방식으로서, 이용자가 입력한 인구통계학적 정보(나이, 지역 등)를 기반으로 이용자와 콘텐츠의 연관성을 예측한다[1].

1.2.2 콘텐츠 기반 필터링 기법

콘텐츠 기반 필터링 기법은 이용자가 과거에 좋아했던 항목과 유사한 특징들을 가지는 항목을 추천하는 기법으로, 이용자와 콘텐츠 사이에 유사도를 계산해서 연관성이 높은 콘텐츠를 추천하는 기법이다. 이 기법은 이용자와 추천 콘텐츠의 프로필 정보를 모두 알고 있을 때 사용할 수 있다[4]. 관련하여 웹, IPTV 등의 콘텐츠 유통망에서의 개인화 추천서비스를 위하여 이용자의 콘텐츠 이용행위와 콘텐츠의 위치정보를 활용한 추천 방법을 제안하는 연구에서 협업 필터링 기법 및 콘텐츠 기반 필터링 기법을 활용한 바 있다[5].

1.2.3 협업 필터링 기법

협업 필터링 기법은 이용자들의 선호도를 기반으로 유사한 성향을 가지는 다른 이용자들을 식별하고, 식별된 다른 이용자가 가지는 정보 중 이용자가 없는 정보를 추천해 주는 방식으로 현재까지 가장 널리 쓰이고 있다[6]. 이러한 협업 필터링 방식은 빅데이터 기술의 활용 이후 이용자들의 실시간 콘텐츠 평가를 통해 수집되는 콘텐츠 평점을 비롯하여 콘텐츠 구매 또는 이용횟수, 콘텐츠별 이용자 체류시간, 열람횟수, 콘텐츠 열람

경로 등 다양하고 풍부한 로그정보를 분석하여 추천 결과를 생성한다. 협업적 필터링은 메모리 기반(Memory-based) 방식과 모델 기반(Model-based) 방식으로 분류할 수 있다. 이와 관련한 연구는 방송, SNS, IoT 등 다양한 분야에서 이루어졌다. 방송 분야에서는 방송콘텐츠를 소비한 이용자의 소비이력 정보를 바탕으로, 협업 필터링 알고리즘을 사용하여 콘텐츠간의 유사도를 도출하여 추천을 제공하는 연구가 수행되었다[7]. SNS를 대상으로 한 연구는 소셜 네트워크 환경에서 협업 필터링 기법을 이용하여 그룹 추천 기법을 제안한 연구[8]와 사용자 성향 및 협업 필터링을 이용하여 이벤트 추천 기법을 제안한 연구[9], 사용자 행위를 고려하여 적합한 콘텐츠를 추천하는 기법에 대한 연구[10]가 수행되었다. IoT 환경에서 그룹 사용자들의 서비스 선호도를 분석하고, 이를 새로운 그룹 사용자에게 서비스를 추천할 수 있는 방법을 개발한 연구[11]에서 그룹의 구성 정보를 기반으로 그룹들 간의 유사도를 측정하여 사용자 기반 협업 필터링을 적용하였다.

본 연구에서는 기본적으로 협업 필터링 기법을 활용하였으나, 신규 콘텐츠의 경우 이용자를 대상으로 한 홍보를 위하여 우선 추천하는 방식을 적용하였다.

1.2.4 하이브리드 필터링 기법

하이브리드 필터링 기법은 콘텐츠 기반 필터링 기법과 협업 필터링 기법의 단점을 극복하기 위해 여러 형태의 다른 추천 시스템을 결합하는 것이다. 그 중 최근 대형 온라인 쇼핑몰 업체인 아마존닷컴에서 이용하고 있는 아이템 간 협업 필터링(Item-to-item collaborative filtering) 기법이 실제 환경에서 성공적으로 적용되고 있는 기법이다[1]. 이 추천 기법은 다른 이용자가 구매한 아이템의 정보를 통해 이용하지 않은 아이템의 점수를 예측하는데, 특히 이용자가 선호하는 아이템과 유사하거나 높은 평점을 줄 것 같은 아이템을 찾기 위해 상관계수나 코사인유사도(Cosine similarity)를 사용해 아이템 간의 유사도를 계산한다[1]. 하이브리드 필터링 기법을 이용하여 영화 추천 알고리즘[12]과 IPTV 콘텐츠 추천 서비스 모델을 제안한 연구[13] 등이 이루어진 바 있다.

2. 빅데이터를 기반으로 한 개인화된 정보 추천

2.1 빅데이터

이용자에게 적절한 추천 서비스를 제공하기 위해서는 이용자가 가지고 있는 특성을 파악하고 해당 사용자 특성과 매치되는 특성을 가진 콘텐츠를 선별해내야 한다[1]. 그러나 모든 콘텐츠에 대한 데이터가 기존에 비해 너무 방대하기 때문에 종래의 데이터 처리 기법으로는 적절한 콘텐츠를 선별할 수 있는 알고리즘을 구현하기가 쉽지 않다. 그래서 최근 빅데이터 처리 기법이 대두되고 있다. 빅데이터 처리 솔루션들은 대용량의 데이터 웨어하우스와 고속의 데이터 분산처리 기법을 이용하여 기존 데이터 분석 시스템이 처리할 수 없는 작업을 할 수 있다. 대표적인 빅데이터 처리 프레임워크는 하둡(Hadoop)이며, 페이스북, 트위터, 야후 등 많은 업체들이 사용하고 있다[14].

2.2 개인화

일반적으로 개인화라는 용어는 이용자의 정보요구에 부합하는 정보를 제공한다는 의미로 광범위하게 사용되고 있다. 개인화는 소극적 개인화와 적극적 개인화로 구분할 수 있다. 소극적 개인화는 이용자의 특정한 정보 검색과정에서 이용자의 프로파일을 기반으로 이용자에게 적합한 정보 및 상품을 제공하는 것으로 주로 정보검색 분야에서 용어의 중의성 해소 등의 내용이 연구되고 있다. 적극적 개인화는 서비스 제공자가 새로운 정보가 입수될 때 요청이 있기 전에 이용자의 선호도를 반영하고 있는 프로파일을 기반으로 적합한 정보를 제공하는 서비스로써 개인화 추천 서비스라고 할 수 있으며, 주로 데이터마이닝 분야에서 활발히 연구되고 있다 [15].

III. 연구 방법

1. 기존 시스템

본 연구에서는 국내의 과학 학술정보 사이트 중 국내 뿐만 아니라 해외에 거주하고 있는 한인 과학자들의 학술 연구 분야에서의 학술 활동 및 동료 연구자와의 학

술적 커뮤니케이션 활성화를 목적으로 하는 플랫폼인 KOSEN을 대상으로 추천 알고리즘 고도화의 전후 비교 평가를 진행하였다. 현재 KOSEN을 이용하고 있는 회원은 134,708명(2017.8.29.기준)으로 해외 회원이 10,623명이고, 석사 학위 이상 소지자는 전체 회원의 59.56%에 이른다. 그리고 해외 회원 중 53.85%가 박사 학위를 소지하고 있다.

KOSEN은 2017년 3월부터 선제적 개인화 서비스를 제공하기 위한 개인별 맞춤 서비스를 제공하고 있다. 이는 회원들이 활동한 로그를 기반으로, 빅데이터 처리 기술에 기반하여 최근에 이용한 콘텐츠간의 유사도를 기반으로 추천 콘텐츠를 추출해 개인화 서비스를 제공한다. 회원에게 적합한 서비스 및 콘텐츠를 필요한 시점에 제공하기 위하여 서비스 메뉴와 각 메뉴별 추천 콘텐츠를 회원 그룹별, 개인별로 생성하여 주기적으로 활동 분석 결과를 시스템에 자동 반영하고 있다. 회원의 활동로그가 있는 경우, 이용했던 콘텐츠와 유사한 콘텐츠를 노출하고, 신규회원과 같이 활동로그가 없는 경우, 기존 회원 중 신규회원과 유사한 회원에 대한 활동로그를 기반으로 콘텐츠를 제공한다. KOSEN의 이용자는 크게 회원과 비회원으로 구분하고, 회원은 다시 보통회원, 휴면회원, 신규회원으로 구분한다. 초기 사용 정보가 거의 없는 휴면회원과 신규회원은 휴면회원과 신규회원이 속해져 있는 유저 그룹의 서비스를 추천하여 사용을 유도하고, 비회원에게는 관리자가 설정한 정보를 제공한다.

- 회원
 - 보통회원 : 최근 180일 이내에 로그인 이력이 있는 회원
 - 휴면회원 : 최근 180일 이내에 로그인 이력이 없는 회원
 - 신규회원 : 최근 60일 이내에 가입한 회원
- 비회원

KOSEN에서 운영하고 있는 '서비스 메뉴 및 콘텐츠 개인화 추천 시스템'의 개요는 다음과 같다.

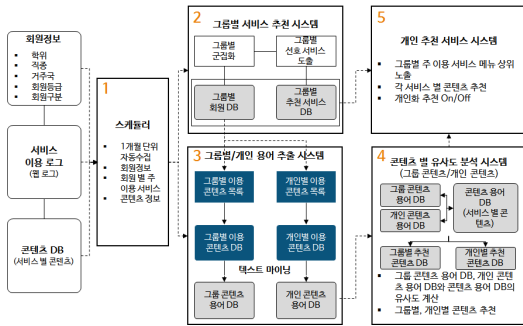


그림 1. '서비스 메뉴 및 콘텐츠 개인화 추천 시스템'의 개요

1.1 스케줄러

본 시스템은 개인을 식별하여 유사한 성격을 보유하고 있는 이용자를 군집화하고, 그를 통해 구분된 이용자 그룹의 선호 서비스를 도출하기 위해 개인 이용자 프로파일 및 서비스 이용로그를 주기적으로 수집한다. 수집 주기는 1개월 단위이며, 서비스 선호도, 키워드, 분야, 회원 활동로그, 시간, 아이템기반 추천, 유저기반 추천 등의 요소를 고려한다.

1.2 그룹별 서비스 추천 시스템

회원정보 및 회원의 활동정보를 기반으로 데이터를 수집하며, 수집된 데이터를 R 프로그램의 군집화(Clustering) 기능을 이용하여 자동 클러스터링 하는데, 이 때 K-means 알고리즘을 이용한다. 그리고 회원의 특성을 구분하고 클러스터링 시 유저그룹의 변별력을 확연히 구분할 수 있도록 '학위, 직종, 거주국, 회원등급, 회원구분'에 가중치를 적용하여 유사한 성격의 이용자 그룹 9개를 도출하였다. 도출된 그룹별 회원정보를 자동 저장하며, 매월 단위로 이용자의 특성정보를 기반으로 그룹을 재구성하여 DB에 자동 저장한다. 그리고 각 그룹별로 주로 이용하는 서비스와 서비스 이용로그를 Mapping 처리하여 그룹별 선호 서비스를 분석하여 추천하고, 추천된 각 그룹별 서비스 정보를 DB에 자동 저장 처리하여 그룹별 회원 DB정보와 그룹별 추천 서비스 DB를 구성한다.

1.3 그룹개인별 용어 추출 시스템

그룹·개인별 용어 추출 시스템은 그룹, 개인이라는

관점만 다를 뿐 동일한 절차로 구성된다. 각 그룹별 회원 DB와 수집된 서비스 이용로그 중 각 그룹별 사용자 ID 정보와 콘텐츠 정보를 이용한 사용자 ID를 Mapping 처리하여, 각 그룹별 이용자가 이용한 전체 콘텐츠 목록을 생성한다. 그리고 그룹별 콘텐츠 목록 정보 중 각 콘텐츠의 제목, 키워드, 초록 정보를 R 프로그램의 자연어 분석 패키지인 KoNLP 라이브러리를 이용하여 텍스트마이닝을 통해 텍스트 정보를 추출한다. 그렇게 각 그룹별 콘텐츠에서 추출된 텍스트 정보를 하나의 문서 번호(Document ID)와 다수의 단어로 구분해 관리한다. 또한 문서번호에 부착된 각각의 텍스트 정보 중 유의미한 단어의 생성에 저해되는 정보(조사, 숫자, 특수문자, 원문자, 스크립트 문자 및 더블스페이스 등)를 필터링(Filtering out)하여, 개별 이용자가 유사한 그룹 내에서 이용한 콘텐츠의 용어를 분석하여 각 그룹·개인별 콘텐츠 용어 DB를 구축한다. 즉, 개별 이용자와 유사한 그룹 내에서 이용한 콘텐츠의 용어를 분석하여 그룹 콘텐츠 용어 DB를 구비하고, 개별 이용자가 이용한 콘텐츠의 용어를 분석하여 개인 콘텐츠 용어 DB를 구비하는 것이다.

1.4 콘텐츠 별 유사도 분석 시스템

각 그룹·개인별 콘텐츠 용어와 전체 콘텐츠의 용어 DB(DTM : Document Term Matrix)의 유사도를 TF-IDF(Term Frequency-Inverse Document Frequency) 기법을 통해 계산한다. TF-IDF 기법은 특정한 단어가 문서 내에서 얼마나 자주 등장하는지를 나타내는 값을 의미하며, 이 값이 높을수록 문서 내에서 중요한 단어(키워드)라고 생각할 수 있다. 다만 해당 단어가 다른 문서 내에서도 많이 나온다면 중요하지 않은 단어일 수 있다. TF-IDF 기법을 통해 계산된 각 그룹·개인별, 서비스 별 콘텐츠의 상관관계 값이 높은 순서에 따라 Ranking을 구성하여 DB에 저장하고 그룹·개인별 추천 DB를 생성한다.

1.5 개인 추천 서비스 시스템

그룹별 서비스 추천 시스템에서 분석된 그룹별 추천 서비스 DB와 콘텐츠별 유사도 분석 시스템을 통해 구

비된 그룹별 콘텐츠 DB, 개인별 콘텐츠 DB를 기반으로 개인 추천이 이루어진다.

1.5.1 서비스 추천

시스템에 로그인 한 이용자를 기준으로 이용자가 편제된 그룹군을 자동 체크하여 이용자가 속한 그룹군이 주로 이용하는 서비스 메뉴의 Ranking을 통해 순서대로 서비스 메뉴를 자동 구성하여 제공함으로써 이용자의 정보 확보에 대한 시간적 제약을 해소할 수 있다.

1.5.2 콘텐츠 추천

각 서비스 메뉴의 콘텐츠 이용 내역이 있는 경우, 콘텐츠 별 유사도 분석 시스템에서 구비된 개인별 추천 콘텐츠 DB를 활용하여 이용자 기반의 추천 콘텐츠를 제공하고, 이용자가 특정 서비스 메뉴에 대한 사용 이력이 없는 경우, 개인 이용자가 포함되어진 그룹에서 추천된 그룹 콘텐츠 정보를 개인의 연구 분야에 맞추어 제공한다. 그리고 이용자는 로그인 이후, 서비스 추천 및 콘텐츠 추천기능을 “On 또는 Off” 할 수 있다. 로그인 이후라도 서비스 추천 및 콘텐츠 추천의 기능 Off를 통해 최신 등록일자 기준의 콘텐츠를 제공 받을 수 있다.

2. 콘텐츠 추천 알고리즘 개선

기 운영 중인 콘텐츠 추천 알고리즘은 KOSEN의 회원정보 및 회원의 활동정보를 기반으로 데이터를 수집하고, 수집된 데이터를 ‘학위, 직종, 거주국, 회원등급, 회원구분’에 가중치를 적용하여 자동 클러스터링을 통해 이용자 그룹 9개를 도출하였다. 본 연구에서는 이용자의 추천 정확도를 높이기 위해 이용자의 ‘전공’ 항목의 추가 적용을 통해 개인별 콘텐츠 추천 알고리즘을 최적화하고, 기존 알고리즘과 최적화된 알고리즘을 통한 추천 결과의 성능평가를 수행하였다.

즉, ‘학위, 직종, 거주국, 회원구분, 회원등급’의 총 5개 정보를 추출하고, 클러스터링 시 회원의 ‘전공’ 항목을 포함하여 9개 그룹을 분석하였다. 또한 추천 콘텐츠 선정 시, 개인의 전공 분야와 콘텐츠의 전공 분야 간 유사도 계산을 통해 추천 콘텐츠를 생성하였다.

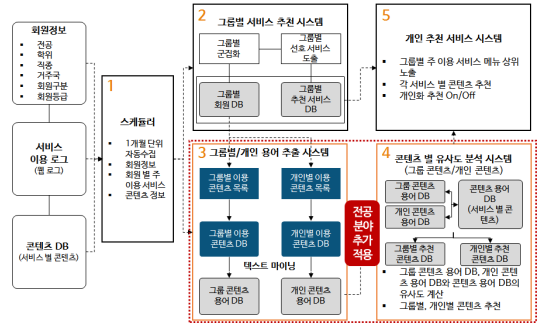


그림 2. ‘서비스 메뉴 및 콘텐츠 개인화 추천 시스템’에 전공 분야 추가 적용

3. KOSEN 서비스 특성 구분

KOSEN 개인화 메인 페이지에서 제공되는 서비스는 크게 리스트식 서비스와 카드식 서비스로 나눌 수 있다. 리스트식 서비스에는 학회정보, 지식공감, What is?, 코센리포트, 분석신청, 동향보고서, 연구실정보 등의 서비스가 제공되고, 카드식 서비스에는 동영상자료, 지식큐레이터, 글로벌뉴스, 이슈토론 등의 서비스가 제공된다.

이 중 가장 이용률이 높은 ‘코센리포트’ 서비스에 대한 평가를 진행하였다. 코센리포트는 최신 과학기술동향에 대한 전문가 분석보고서(2015년 422건, 2014년 521건)와 학회보고서(2015년 95건, 2014년 118건)를 포함한다.

4. 성능 평가 진행

최적화된 알고리즘이 실제로 이용자들에게 더 적합한 콘텐츠를 추천하는지 확인하기 위해, 기존에 운영했던 알고리즘으로 추출된 콘텐츠 목록과 최적화된 알고리즘으로 추출된 콘텐츠 목록을 실제 이용자들에게 제시하여 관련성을 평가하게 하였다.

비교평가를 진행할 대상은 다음과 같은 절차로 선정하였다. KOSEN의 회원 등급은 관리자, 교정자, 전문가, 준전문가, 일반회원으로 구분되는데, 이 중 콘텐츠 추천의 변화를 인지하기 용이하도록 전문가, 준전문가, 일반회원에 한정하여 각각 활동로그가 가장 많은 상위 10명, 총 30명을 선정하였다. 이 중 관리자와 관련 있는 2명을 제외하고 28명을 대상으로 평가를 진행하였다.

평가 방법은 다음과 같다. 이용자마다 각각 본인의 정보를 바탕으로 기존에 운영했던 알고리즘으로 추출된 ‘코센리포트’ 서비스의 콘텐츠 목록과 최적화된 알고리즘으로 추출된 ‘코센리포트’ 서비스의 콘텐츠 목록을 각각 15건씩 제시하였다. 이용자는 그 리스트를 보고, 각 항목에 대하여 본인과의 관련성 여부를 평가하였다.

IV. 연구 결과

28명 중 회신된 11명의 응답을 분석하여 결과를 도출하였다. 평가 결과 최적화된 알고리즘으로 추천된 콘텐츠가 상대적으로 본인과의 관련성이 높다고 평가하였다. 상세한 결과는 [표 1]과 같다.

표 1. 알고리즘별 콘텐츠 추천 관련성 평가 결과

분야	회원 구분	회원	콘텐츠관련도(%) ¹⁾		증가율 (%)
			기존 알고리즘	최적화 알고리즘	
보건의료	전문가	이**	73.3	100.0	26.7
		이**	40.0	46.7	6.7
	일반회원	정**	6.7	53.3	46.7
생명과학	전문가	김**	40.0	80.0	40.0
	전문가	심**	73.3	100.0	26.7
에너지·자원	전문가	신**	13.3	13.3	0
		이**	13.3	26.7	13.3
전기전자	준전문가	이**	13.3	13.3	0
		정**	0	26.7	26.7
화학	준전문가	강**	0	20.0	20.0
		김**	33.3	60.0	26.7
	평균		27.9	49.1	21.2

※ 분야별 정렬

[표 1]을 보면 기존 알고리즘에서 도출된 콘텐츠 관련도는 위와 같으며, 그 평균은 27.9%이다. 콘텐츠 관련도는 도출된 전체 콘텐츠개수 15개를 도출된 콘텐츠 중 본인과 관련있다고 선택한 개수로 나누어 비율을 산정하였다. 최적화된 알고리즘에서 도출된 콘텐츠 관련도는 평균 49.1%로 21.2%의 차이를 보였다. 따라서 최적화 알고리즘으로 추천되는 콘텐츠가 회원에게 더 적합한 정보를 추천함을 알 수 있다.

다음은 [표 1] 내용의 이해를 돕기 위하여 그림으로 표현한 것이다.

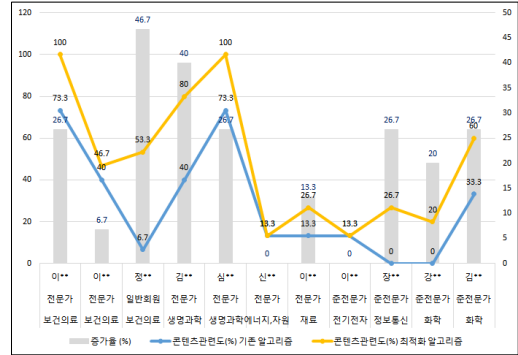


그림 3. 알고리즘별 콘텐츠 추천 관련성 평가 결과

실제 응답 결과를 세부적으로 살펴보면, 화학 분야 중 준전문가인 김** 회원은 기존 알고리즘에서 도출된 콘텐츠 중 본인과 관련 있다고 선택한 개수는 5개였으나, 최적화된 알고리즘에서 도출된 콘텐츠에서는 9개가 본인과 관련있다고 선택하여, 콘텐츠 관련도가 33.3%에서 60%로 증가하였다.

실제적으로 회원에게 제시된 리스트 중 상위 5개를 보면 [표 2][표 3]과 같다. recommend ranking은 번호 순으로 회원과 더 관련도가 높은 콘텐츠를 의미한다. [표 2]에는 기존 알고리즘으로 추출된 콘텐츠가 있으며 다양한 주제 분야 중 물리학 분야의 콘텐츠만 본인과 관련 있는 것으로 평가하였다. 그런데 [표 3]에서는 회원의 전공 분야인 화학 분야의 콘텐츠가 추천되었으며, 그 중 1건은 본인과 관련이 없는 것으로 평가하였다.

이를 통해 콘텐츠 추천 알고리즘에 회원의 전공 분야를 적용함으로써 이용자와 상관없는 자료가 제거되고, 이용자와 관련있고 흥미 있는 자료가 상위에 랭킹되어 이용자에 대한 추천 정확도가 향상되었음을 알 수 있다. 그러나 전공 분야가 추가됨에 따라 전공 분야에 적합한 자료가 상위에 랭킹됨으로써, 기존에 관련성이 있다고 판단했던 물리학 분야의 자료는 추천되지 않는 경우도 발생하게 되었다.

1 산출방법 = $\frac{\text{도출된 콘텐츠 중 본인과 관련있다고 선택한 개수}}{\text{도출된 전체 콘텐츠 개수 (15개)}} \times 100$

표 2. 기존 알고리즘 콘텐츠 추천 관련성 평가 리스트

recommend ranking	title	subject	relation
1	제68회 미국물리학회 유체역학 부문 연례회의(학회보고서)	물리학	○
2	36th Annual Conference on Information Sciences and Systems	전기/전자	×
3	미국 핵무기 계획을 위한 삼중수소 생산: 주요 현안의 분석	지구과학	×
4	AVS 53rd International symposium & exhibition	재료	×
5	2000 AAPS Annual Meeting and Exposition	보건의료	×

표 3. 최적화 알고리즘 콘텐츠 추천 관련성 평가 리스트

recommend ranking	title	subject	relation
1	Fuel Cell Sysyems for Distributed Generation	화학	○
2	제252회 미국화학회 추계학술발표회	화학	○
3	무기화학의 발전방향	화학	○
4	13th Asian Chemical Congress	화학	○
5	선진국의 벤처창업 및 벤처캐피탈 투자증대, 맞춤형 기술인력 양성을 위한 정책 조사	화학	×

V. 결론

본 연구에서는 과학 학술정보 서비스 플랫폼에서 운영 중인 서비스 메뉴와 각 서비스 별 콘텐츠 정보를 제공하는 알고리즘 중 콘텐츠 추천 알고리즘을 최적화하여 그 성능을 비교평가 함으로써, 이용자에 대한 추천 정확도를 높일 수 있는 방안을 제시하고자 하였다.

KOSEN의 개인화 메인 페이지에서 제공되는 리스트 식 서비스와 카드식 서비스 중 가장 이용률이 높은 ‘코센리포트’ 서비스를 대상으로 성능평가를 진행하였으며, 성능평가 결과 이용자의 ‘전공’ 항목을 적용하여 최적화된 개인별 콘텐츠 추천 알고리즘의 효과를 검증하였다. 기존 알고리즘에서 도출된 콘텐츠 관련도의 평균은 27.9%이며, 최적화 알고리즘에서 도출된 콘텐츠 관련도는 평균 49.1%로 21.2%의 차이를 보였다. 이 차이는 콘텐츠 추천 알고리즘에 회원의 ‘전공’ 항목을 적용

함으로써 회원과 상관없는 자료가 제거되고, 이용자와 관련있고 흥미 있는 자료가 상위에 랭킹되어 이용자에게 대한 추천 정확도가 향상되었음을 나타낸다.

이는 이용자에게 적합한 콘텐츠를 시스템에서 자동 도출하여 각 서비스 메뉴 별로 제공함으로써 정보 획득 시간을 단축하고, 연구정보로서 가치 있는 연구결과물의 생명주기를 연장할 수 있는 방안이 될 것이라 기대한다.

다만 본 연구에서 제시한 결과는 ‘코센리포트’ 서비스 한 가지만을 대상으로 평가를 진행하였기 때문에 알고리즘의 성능을 일반화하기에는 한계가 있다. 향후 다른 서비스를 대상으로 개인별 콘텐츠 추천 알고리즘의 성능을 평가할 예정이다. 또한 기존의 알고리즘에 ‘전공’ 분야만을 추가하였기 때문에 다른 변수들도 추가하여 변수들의 활용에 대한 연구를 진행함으로써 콘텐츠 추천 알고리즘을 고도화할 예정이다.

참고 문헌

- [1] 유영석, 김지연, 손방용, 정종진, “온라인 음악 콘텐츠 추천 시스템 구현을 위한 협업 필터링 기법들의 비교 평가,” 전기학회논문지, 제66권, 제7호, pp.1083-1091, 2017.
- [2] 오진오, 유환조, “추천 시스템,” 정보과학회지, 제32권, 제1호, pp.53-58, 2014.
- [3] I. Jeong, X. Yang, and H. Jung, “A Study on Movies Recommendation System of Hybrid Filtering-Based,” Journal of the Korea Institute of Information and Communication Engineering, Vol.19, No.1, pp.113-118, 2015.
- [4] L. Iaquina, M. de Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino, “Introducing serendipity in a content-based recommender system,” Proc. International Conference on Hybrid Intelligent Systems, pp.168-173, 2008.
- [5] 김용, 김문석, 김윤범, 박재홍, “이용자 이용행위 및 콘텐츠 위치정보에 기반한 개인화 추천방법에

- 관한 연구,” 정보관리학회지, 제26권, 제1호, pp.81-105, 2009.
- [6] 안성만, 김인환, 최병구, 조운호, 김은홍, 김명균, “소셜네트워크 분석을 통한 협업필터링 추천 성과의 이해,” 한국전자거래학회지, 제17권, 제2호, pp.129-147, 2012.
- [7] 오수영, 오연희, 한성희, 김희정, “사용자 소비이력기반 방송 콘텐츠 추천 시스템,” 방송공학회논문지, 제17권, 제1호, pp.129-139, 2012.
- [8] 양희태, 차재홍, 안민제, 임종태, 이하, 복경수, 유재수, “동적 사용자 프로파일 및 협업 필터링을 이용한 소셜 네트워크 그룹 추천,” 한국콘텐츠학회논문지, 제13권, 제11호, pp.11-20, 2013.
- [9] 복경수, 이수지, 노연우, 김민수, 김연우, 임종태, 유재수, “소셜 네트워크에서 사용자 성향 및 협업 필터링을 이용한 이벤트 추천 기법,” 정보과학회 컴퓨터링의 실제 논문지, 제22권, 제10호, pp.504-512, 2016.
- [10] 고건식, 김병훈, 김대운, 최민웅, 임종태, 복경수, 유재수, “소셜 네트워크 환경에서 사용자 행위를 고려한 콘텐츠 추천 기법,” 한국콘텐츠학회논문지, 제17권, 제2호, pp.404-414, 2017.
- [11] 이진서, 고인영, “사물 인터넷 환경에서의 그룹 사용자를 위한 그룹 구성 정보 기반 서비스 추천 방법,” 정보과학회논문지, 제43권, 제7호, pp.786-794, 2016.
- [12] 김상화, 오병화, 김문중, 양지훈, “협력적 필터링과 콘텐츠 정보를 결합한 영화 추천 알고리즘,” 정보과학회논문지 : 소프트웨어 및 응용, 제39권, 제4호, pp.261-268, 2012.
- [13] 조도는, 김시정, 곽윤식, “사용자 선호도 학습을 이용한 개인화 콘텐츠 추천 방법 연구,” 한국정보기술학회논문지, 제9권, 제9호, pp.229-235, 2011.
- [14] 최성우, 한성희, 정병희, “협업 필터링 기반의 콘텐츠 추천 시스템과 빅데이터 처리 솔루션을 이용한 상용화 개발 방향,” 방송과 미디어, 제19권, 제4호, pp.50-59, 2014.
- [15] 김용, “Apriori 알고리즘 기반의 개인화 정보 추

천 시스템 설계 및 구현에 관한 연구,” 한국비블리아학회지, 제23권, 제4호, pp.283-308, 2012.

저 자 소 개

박 성 은(Seong-Eun Park)

정회원



- 2014년 8월 : 성균관대학교 문헌정보학과(석사)
- 2016년 7월 ~ 현재 : 한국과학기술정보연구원, 정보융합연구실

<관심분야> : 과학 학술정보 플랫폼, 콘텐츠 추천

황 윤 영(Yun-Young Hwang)

정회원



- 2004년 2월 : 충남대학교 컴퓨터공학(석사)
- 2011년 2월 : 충남대학교 컴퓨터공학(박사)
- 2012년 9월 ~ 현재 : 한국과학기술정보연구원, 정보융합연구실

<관심분야> : 정보표준화, 개인화 적용 콘텐츠 관리

윤 정 선(Jungsun Yoon)

정회원



- 1991년 2월 : 카이스트 전산학과(공학사)
- 1993년 2월 : 카이스트 전산학과(공학석사)
- 2000년 5월 ~ 현재 : 한국과학기술정보연구원, 과학기술정보센터

<관심분야> : 추천 알고리즘, 개인화 서비스