필터링 기법을 이용한 도서 추천 시스템 구축

Developing a Book Recommendation System Using Filtering Techniques

정 영 미*· 이 용 구**

Young-Mee Chung· Yong-Gu Lee

차 례

1. 서 론
2. 추천기법
3. 도서 추천 기법 실험
4. 실험 결과 분석
5. 결 론
· 참고문헌

초 록

이 연구에서는 최근에 주목받고 있는 협업 필터링 기법을 중심으로 여러 가지 추천 기법을 살펴본 후 대출대상 도서의 추천 시스템을 구축하였다. 연관성 규칙 기반 기법, 협업 필터링 기법, 내용기반 필터링 기법을 응용하여 실제 대학도서관에서 특정 이용자가 대출할 만한 도서를 추천하는 시스템을 구현하고 각 기법의 추천 성능을 평가하였다. 실험 결과 대출대상 도서를 추천하는 데 있어 협업 필터링 기법과 내용기반 필터링 기법을 각각 따로 적용하는 것보다 두 기법을 함께 이용한 혼합형 필터링 추천 기법이 더욱 효과적인 것으로 나타났다.

키 워 드

도서 추천 시스템, 추천 기법, 협업 필터링, 내용기반 필터링, 연관성 규칙

* 연세대학교 문헌정보학과 교수
(Professor, Library & Information Science Dept., Yonsei University, ymchung@yonsei.ac.kr)
** 연세대학교 문헌정보학과 대학원
(Graduate School of Library & Information Science, Yonsei University, yonggulee@hotmail.com)

ABSTRACT

This study examined several recommendation techniques to construct an effective book recommender system in a library. Experiments revealed that a hybrid recommendation technique is more effective than either collaborative filtering or content-based filtering technique in recommending books to be borrowed in an academic library setting. The recommendation technique based on association rule turned out the lowest in performance.

KEYWORDS

Book Recommendation System, Recommender System, Recommendation Techniques, Collaborative Filtering, Content-based Filtering, Association Rule

1. 서 론

최근 정보기술의 발전에 따라 이용자가 접근할 수 있는 정보의 양은 기하급수적으로 늘어나고 있으며, 다양한 정보 매체에 의한 정보의 범람은 이미 개인이 처리할 수 있는 한계를 넘어섰다. 이러한 정보과잉 속에서 이용자가 원하는 정보를 단시간 내에 검색해내는 일은 점점 어려워지고 있다. 이미 보편화되어 있는 도서관의 SDI (Selective Dissemination of Information) 서비스나 웹 기반 정보필터링 시스템은 이용자가 원하는 정보만을 제공해 줌으로써 정보과잉의 문제를 해결하는 도구로 사용되고 있다. 최근에는 필터링 기법을 응용한 추천 시스템(recommender system)이 등장하였으며, 이 시스템을 이용하여 이용자가 원하는 정보를 예측하고 추천함으로써 정보과잉의 문제를 해결하려는 노력이 진행되고 있다.

추천 시스템은 정보검색 시스템이나 정보필터링 시스템의 일부분으로 구축된다. 추천 시스템은 이용자가 어떤 정보에 관심이 있는지, 어떤 정보가 유용한지를 예측하는 일에 초점을 맞춘다. 추천 시스템의 예측은 이용자 개인의 프로파일을 이용함으로써 개인화 된다. 이러한 추천 시스템은 초기에는 주로 내용기반 필터링 기법(content-based filtering technique)을 이용하였으나 최근에는 자동화된 협업 필터링 기법(automated collaborative filtering technique)이 많이 이용되고 있다.

자동화된 협업 필터링은 내용기반 정보필터링 시스템을 보완하는 기법으로 최근 각광을 받고 있다. 자동화된 협업 필터링은 인터넷에서 상당한 성공을 보여주었으며 인터넷 상의 최대 서점인 Amazon.com과 최대 CD 상점인 CDNow.com, 그리고 가장 자주 방문되는 영화 사이트 중의 하나인 MovieFinder.com과 같은 사이트에서 이용되고 있다(Herlocker et al. 1999).

정보과잉 문제를 해결하기 위한 추천

시스템을 도서관에도 적용할 수 있다. SDI와 같은 기존의 필터링 서비스는 주로 문헌의 내용을 분석하여 정보를 걸러내는 내용기반 필터링 기법을 이용한다. 이러한 내용기반 필터링 기법과 최근 주목받고 있는 협업 필터링 기법을 같이 적용한다면, 두 기법이 상호 보완을 통해 보다 효과적으로 정보과잉의 문제를 해결할 수 있을 것으로 보인다.

따라서 이 연구에서는 실제 도서관의 소장도서의 대출정보와 서지정보를 이용하여 이용자 관심 분야의 도서를 추천하는 도서 추천 시스템을 구현하고, 실험을 통해 여러 추천 기법 가운데 최적의 추천 결과를 제공하는 추천 기법을 제안하고자 한다.

2. 추천 기법

2.1 연관성 규칙 기반 기법

고객에게 제품을 추천하기 위해 전자상거래에서 자주 사용되는 기법 중의 하나가 연관성 정보를 이용하는 것이다. 이 기법은 특정 트랜잭션에 하나의 제품이 존재하고 동시에 같은 트랜잭션에 다른 제품이 존재할 때 이러한 두 제품 사이의 연관성을 발견하는 데 초점이 맞추어져 있다. 즉 m 제품의 집합 $\{P_1, P_2, \ldots, P_m\}$을 P로 나타내고, 트랜잭션 $T \subseteq P$를 함께 구매한 제품의 집합으로 정의할 때, 여기서 $X, Y \subseteq P$이고 $X \cap Y = 0$ 인 제품 집합 X와 Y 사이의 연관성 규칙은, 트랜잭션 T에서 집합 X의 제품이 존재한다면 또한 집합 Y의 제품이 동시에 트랜잭션 T에 존재한다는 강한 확률을 나타낸다(Sarwar et al. 2000).

구매 데이터에서 연관성 규칙을 이용하여 추천하는 대표적인 사례는 세계 최대의 온라인 서점인 Amazon.com에서 찾아 볼 수 있다. Amazon.com의 경우 수많은 도서에 대한 다양한 정보를 수록한 웹 페이지를 제공하고 있으며, 특히 'Customers Who Bought' 항목은 구매자들이 공통적으로 구매한 도서에 연관성 규칙을 적용하여 생성한 추천 정보를 제공하고 있다(Schafer, Konstan, and Riedl 1999).

2.2 협업 필터링 기법

협업 필터링 기법은 추천 시스템에서 가장 흔히 쓰이는 기법이다. 협업 필터링 기법을 이용한 추천 시스템은 이용자의 선호도를 수집하여 데이터베이스를 구축하고, 특정 이용자와 유사한 취향이나 정보요구를 갖는 이웃들을 데이터베이스에서 찾아내어, 이들이 좋아하는 물건이나 정보를 이용자에게 추천한다.

협업 필터링이란 용어는 1990년 초에 제록스 PARC 연구소에서 개발한 Tapestry에서 처음으로 사용되었으며(Goldberg et al. 1992), 자동화된 협업 필터링 기법은 유즈넷 뉴스의 과잉문제를 해결하기 위해 구축된 초기 추천 시스템 GroupLens에서 구현되었다(Resnick et al. 1994). 또한 비슷한 시기에 음악 추천 시스템인 Ringo (Shardanand, and Maes 1995)와 비디오 추천 시스템인 Bellcore(Hill et al. 1995)가 출

현하였다.

일반적으로 협업 필터링 알고리즘에서 아이템을 추천하기 위한 과정은 표현 작업(representation), 이웃 형성(neighbor formation), 추천 생성(recommendation generation)위 세 단계로 나누어진다. 표현 작업은 이용자가 평가했던 아이템을 이용하여 모형화 하는 단계이다. 이웃 형성은 유사한 이웃을 식별하기 위한 단계이며, 추천 생성은 이웃하는 고객으로부터 상위 N개의 추천 아이템을 찾아 이용자에게 제공하는 단계이다.

(1) 표현 작업

전형적인 협업 필터링 기반 추천 시스템에서 입력 데이터는 아이템 m 개와 이용자 n 명의 선호도 집합이다. 이것은 대개 m×n(이용자-아이템) 행렬 R로 표현된다. 즉 r_{ij}는 만약 i번째 이용자가 j번째 아이템을 선호한다면 1이고 그렇지 않으면 0이다. 물론 선호도의 값은 평가 척도에 따라 다르다. 즉 5점 척도를 이용하여 선호도를 평가할 경우 1부터 5의 값이 들어간다.

이용자-아이템 행렬에서 선호도를 입력하기 위한 방식은 이용자가 의식적으로 아이템에 대해 자신의 선호도를 표현하는 명시적 평가(explicit ratings)와 이용자의 행태나 선택을 시스템이 해석하여 평가하거나 선호도의 값을 부여하는 암묵적 평가(implicit ratings)로 구분된다(Breese, Heckerman, and Kadie 1998).

(2) 이웃 형성

협업 필터링 기법을 이용하는 추천 시스템에서 가장 중요한 단계는 이용자 사이의 유사도를 계산하는 일이다. 이러한 일을 수행하기 위해 먼저 특정 이용자와 유사한 취향을 가진 이용자 사이의 이웃집단을 형성해야 한다.

추천 시스템의 알고리즘은 이웃형성 과정을 통해 모형을 만들거나 학습하는 과정이다. 이웃형성의 주 목적은 각각의 이용자 u에 대해 순위화된 이용자 리스트 l을 찾는 것이다. 즉 이용자 집단 $N = \{N_1, N_2, \ldots, N_n\}$이고 이용자 $u \notin N$일 때, 유사도 $sim(u, N_1)$이 최대값, 유사도 $sim(u, N_2)$가 다음 최대값인 순위화된 이용자들을 찾아내는 것이다.

따라서 이웃집단을 형성하기 위해서는 두 가지 단계를 거쳐야 한다. 먼저 특정 이용자와 모든 다른 이용자 사이의 유사도를 구한다. 흔히 사용되는 유사도 가중치 측정 방법은 피어슨 유사계수, 스피어만 순위 유사계수(Spearman rank correlation coefficient), 벡터 유사도(vector similarity) 등이 있다. 그 다음으로 이웃 집단의 규모를 결정한다. 즉 모든 이용자에 대해 계산된 유사도를 가지고 추천 아이템을 예측하기 위해 몇 명의 이웃을 이용할지 결정한다. 이웃집단의 규모를 결정하는 방법은 유사도 가중치가 어느 정도 이상인 모든 이용자를 선택하거나 가장 유사도 가중치가 높은 몇 명의 이용자를 선택하는 방법이 있다.

(3) 추천 생성

협업 필터링 기법을 이용한 추천 시스템의 마지막 단계는 특정 이용자에게 다른

이용자가 평가한 아이템을 추천하는 것이다. 추천을 생성하기 위해 전 단계에서 형성된 특정 이용자의 이웃집단으로부터 상위 N개의 추천 아이템을 추출해서 순위화 할 수 있다. 순위화를 위한 방법은 선호도를 입력하는 척도에 따라 달리 할 수 있다. 즉 선호도의 척도가 구매횟수나 대출횟수와 같은 단순빈도라면 이 단순빈도를 이용하여 순위화 할 수 있다. 그렇지 않으면 추천할 아이템에 대한 특정 이용자의 예측 선호도 값의 정렬을 통해 순위화 할 수 있다. 예측 선호도를 구하는 방법은 특정 아이템에 대한 모든 이용자의 선호도를 가중평균하여 계산하는 기본적인 방법이 있다 (Shardanand, and Maes 1995).

다른 추천 기법과 비교할 때 협업 필터링 기법은 다음과 같은 장점을 갖고 있다 (Burke 2001; Herlocker et al. 1999; Melville, Mooney, and Nagarajan 2001; Funakoshi, and Ohguro 2000).

첫째, 내용이 어떤 형태로 되어있든 정보를 검색할 수 있다. 즉 협업 필터링의 장점은 추천할 대상의 표현에 의존할 필요가 없다는 점이다. 따라서 취향에 따라 다양한 선호도를 보이는 음악이나 영화 같은 복합적인 아이템에도 잘 작동한다.

둘째, 협업 필터링 기법은 아이템에 대한 이용자의 선호도를 수집하고, 같은 취향이나 정보요구를 갖는 이용자들의 선호도와 의견을 바탕으로 특정 이용자의 기호에 맞는 아이템을 추천함으로써 개인화된 추천을 제공할 수 있다.

셋째, 이용자가 검색하고자 하는 내용을 포함하지 않지만 이용자에게 가치 있는 아이템을 추천할 수 있다. 즉 이용자가 원하는 내용이 아이템에는 포함되어 있지 않아도, 유사한 취향의 이용자들에게 좋은 평가를 얻었다면 협업 필터링 기법은 그 아이템을 추천할 수 있다. 왜냐하면 협업 필터링 기법은 질적인 면과 취향에 기반하여 아이템을 여과하는 능력이 있기 때문이다.

반면 협업 필터링 기법의 단점은 다음과 같다.

첫째, 새로운 아이템이 추가되면 누군가가 선호도를 입력하거나 이용하기 전에 그 아이템을 추천할 수 없는 초기 평가자 문제(early-rater problem)가 있다.

둘째, 희박함(sparsity)의 문제이다. 대부분의 영역에서 다수의 아이템에 대해 모든 이용자가 평가할 수 없기에 이용자-아이템 행렬의 평가 값들은 매우 희박한 상태가 된다. 이러한 문제는 다수의 이용자가 평가한 아이템을 찾기가 어려워 협업 필터링 기법을 통한 예측을 힘들게 한다.

2.3 내용기반 필터링 기법

내용기반 필터링 기법은 정보 검색이나 정보 필터링 연구에서 자연적으로 발전하였다. 따라서 내용기반 필터링 기법은 대부분 아이템을 추천하기 위하여 아이템의 내용과 이용자의 정보요구간의 유사도를 측정하고, 그 결과를 순위화 하여 보여준다. 이렇게 아이템의 내용을 중심으로 분석하여 이용자에게 추천하는 기법을 내용기반 필터링 기법이라 한다. 이 기법은 정보검색에 기반을 두고 있으므로 이 분야에서 자

주 사용되는 기법인 가중치 기법, 적합성 피드백, 불리안 검색, 확률검색 모형 등을 활용한다(Balabanovic, and Shoham 1997; Burke 2001).

내용기반 필터링 기법의 단점으로는 영화나 음악 같은 분야에서는 내용 분석이 피상적으로 이루어질 수 있다는 것이다. 현재의 기술로 이와 같은 분야에서 아이템으로부터 유용한 특징을 추출하는 것은 쉽지 않기 때문이다. 또한 내용기반 필터링 기법은 이용자의 프로파일과 비교하여 높은 점수를 얻은 아이템을 추천하므로, 이용자가 이미 평가한 아이템과 유사한 아이템만을 제공하여 과도하게 특수화되는 경향을 보인다(Balabanovic, and Shoham 1997).

2.4 혼합형 필터링 기법

정보 필터링 도구를 발전시키기 위한 관점에서 보면 협업 필터링에 대한 가능성은 매우 크다. 그러나 그 가능성을 극대화하기 위해서는 반드시 내용기반 정보 필터링 기술과 결합될 필요가 있다. 협업 필터링 자체만으로 이용자의 관심이나 취향에 맞는 아이템에 대한 예측은 잘 수행할 수 있으나 특정 아이템의 내용에 대한 이용자의 정보요구를 해결하는 것은 효율적이지 못하다(Herlocker et al. 1999). 내용기반 필터링 기법과 협업 필터링 기법이 각각의 상호보완적인 장단점을 가지고 있기 때문에 보다 정확한 예측을 하기 위해서 두 가지 기법을 결합시키는 연구들이 최근 들어 늘어나고 있다.

3. 도서 추천 기법 실험

3.1 실험 데이터

이 연구의 실험 데이터는 Y대학교 중앙도서관의 단행본 서지정보와 이용자 정보, 그리고 대출· 반납 관련 정보 등으로 구성되어 있다. 데이터 수집 기간은 2000년 9월부터 2001년 8월까지 1년이며, 수집 대상 이용자는 전문화된 주제 분야를 연구하는 석· 박사과정의 대학원생으로 제한하였다. 전체 소장 단행본 826,05권 중에 데이터 수집기간 동안 대출된 단행본은 93,63권이고, 한번이라도 도서를 대출한 학생은 6,466명이다. 대출자수와 주제분야를 고려하여 30개의 질의 문헌을 임의로 추출하였으며, 각 문헌의 주요 서지정보는 <표 1>과 같다.

3.2 실험 내용

도서 추천 실험에서는 특정 문헌과 가장 유사한 문헌을 추천하고 그 결과를 평가하여 각 기법의 성능을 평가한다. 따라서 특정 문헌을 질의로 간주하고 이러한 질의 문헌을 체계적이고 과학적인 방법을 통해 실험 문헌집단으로부터 추출할 필요성이 있다. 일반적인 검색 실험의 경우 실험용 탐색문을 작성하고 이에 대한 검색 성능을 평가하지만 이 실험에서는 이미 이용자의 정보요구를 만족시키는 하나의 문헌을 검색하였다고 전제하고, 이를 이용하여 유사한 다른 문헌을 추천하는 방법을 사용한다. 따라서 추천 성능을 평가하기 위해 기준이

<표 1> 실험용 질의 문헌 30개

번호	서 명	저 자 명	출 판 사	출판년	분류번호
1	Logan의 유한요소법 첫걸음 : ANSYS 예제집	Logan, Daryl L	시그마프레스	1998	620.001515353
2	(실례로 배우는) 시계열분석의 기초 : RATS를 이용한 ARIMA모형의 분석	이윤복	학문사	1999	519.232
3	(Special edition) using 윈도우 2000 professional	Cowart, Robert	인포북	2000	005.4469
4	(Visual C++를 이용한) 디지털 영상처리의 구현	장동혁	PC어드밴스	1999	621.367
5	신경망 이론과 응용, 2	김대수(金大洙)	하이테크정보	1993	006.3
6	계량경제학	이홍윤(李洪閏)	경세원	2000	330.028
7	브랜드 자산의 전략적 관리	Aaker, David A	나남	1992	658.827
8	경험적 가족치료 : Satir이론과 기법	송성자(宋聖子)	中央適性出版社	1994	616.8915
9	구성주의	Schmidt, Siegfried J	까치	1995	149
10	디지털은 자유다 : 인터넷과 지적 재산권의 충돌	홍성태	이후	2000	346.048
11	질적연구 : 방법과 사례	조용환(趙鏞煥)	교육과학사	1999	300.72
12	The elements of real analysis	Bartle, Robert Gardner	Wiley	1976	515.8
13	로마인 이야기, 2 : 한니발 전쟁	염야칠생(鹽野七生)	한길사	1995	937.06
14	의료보장의 기초이론	Normand, Charles E. M	한울	1996	368.42
15	데이터 웨어하우징과 OLAP	조재희(趙載熙)	대청	1996	658.4038
16	주역 왕필주	왕필(王弼)	길	1998	181.41
17	刑法總論	이정원(李廷元)	法志社	1999	345
18	인지적 행동수정의 통합적 접근	Meichenbaum, Donald	양서원	1995	616.8914
19	느낌과 통찰의 음악교육 : 음악교육의 방법	이홍수(李弘守)	세광음악출판사	1992	780.7
20	현대 소설의 이론	Schramke, Ju@rgen	文藝出版社	1995	808.3
21	한국사, 6, 중세사회의 성립, 2	강만길 (姜萬吉)	한길사	1994	951
22	Modern information retrieval	Baeza-Yates, R	Addison-Wesley Longman	1999	025.04
23	한국교육의 사회학적 진단과 처방	김경동(金璟東)	집문당	1999	370.951
24	생태학 : 환경의 위기와 우리의 미래	Odum, Eugene Pleasants	사이언스북스	1995	577
25	우리나라 기업지배구조의 새로운 패러다임 모색	한국 경제연구원	한국경제연구원	1999	658.114
26	티치 유어셀프 XML	Eddy, Sandra E	비앤비	2000	005.72
27	컬러로 보는 생활풍수와 인테리어	Lin, Yun	동도원	1999	133.323
28	Virtual culture : identity and communication in cybersociety	Jones, Steve	Sage Publications	1997	301.243
29	SAS 다변량 통계분석 : 실무자를 위한 실용통계 지침서	유성모	自由아카데미	1999	519.535
30	華西學派의 思想과 民族運動	오영섭 (吳瑛燮)	國學資料院	1999	951.59

되는 질의 문헌을 시스템에 제공하고, 추천 결과인 상위 10개의 순위화된 문헌에 대해 적합성 판정을 하였다.

협업 필터링 추천 기법을 도서관 시스템에 적용하기 위해서는 이용자의 선호도 정보가 필요하다. 추천 대상인 도서에 대해 이러한 선호도를 대체하기 위한 방안으로 도서관 시스템에서 활용할 수 있는 데이터는 대출· 반납 정보에서 찾을 수 있다. 이 실험에서는 특정 문헌이 대출되었다는 것

은 이용자가 그 문헌에 대해 관심이 있다, 즉 선호한다는 뜻으로 해석하였고, 그 선호하는 정도는 그 도서를 얼마나 이용하였는가를 나타내는 대출일수로 측정하였다. 대출일수는 개인별 이용 행태에 따라 달라지므로 선호도 정보로 사용하는 것은 무리가 따르지만 이용자가 명시적으로 선호도를 입력하지 않는 한 도서관 시스템에서 이용 가능한 다른 선호도 정보를 찾기 힘들다. 따라서 이러한 대출일수가 갖는 제한점을 극복하고 개인별 이용 행태를 반영하기 위해 대출일수에서 개인별 평균 대출일수를 빼준 값을 사용하였다.

이 연구에서는 여러 분야에서 활용되고 있는 추천 기법을 도서관 시스템에 응용하기 위해 다음과 같이 세 가지 추천 기법과 혼합형 기법을 도서 추천에 사용하고 각 기법의 성능을 평가하였다.

(1) 문헌간 연관성을 이용한 도서 추천

특정 이용자가 자신의 정보 요구를 해결하기 위해 문헌이나 자료를 이용할 때 같은 주제 분야에 해당하는 여러 권의 도서를 동시에 대출하므로 이러한 대출 패턴을 이용하여 연관성 규칙을 적용하면 이용자에게 특정 문헌을 추천할 수 있다. 다시 말하면 대출정보를 이용한 연관성 규칙 기반 추천 기법은 이용자가 정보요구를 해결하기 위해 도서관에서 대출한 여러 도서간의 연관성을 분석하여, 유사한 정보요구를 갖는 다른 이용자에게 이러한 연관성이 큰 도서들을 추천한다. 이 기법의 실험에서는 연관성을 분석하기 위한 방법으로 대출빈도와 대출자수를 이용하였다.

먼저 대출빈도를 이용한 연관성 측정은 질의문헌 d를 대출한 이용자 집단 $U_d = \{u_{1d}, u_{2d} \cdots u_{nd}\}$와 이들이 대출한 문헌집단 $D_k = \{k_1, k_2 \cdots k_n\}$를 검색한다. 여기서 k_i는 i번째 문헌의 대출빈도로 U_d의 이용자들이 대출한 총 횟수가 되며, 이 값이 d와의 연관성 정도를 나타낸다. 또한 대출자수를 이용한 연관성 측정은 대출빈도를 이용한 연관성 측정 방법과 마찬가지로 문헌집단 $D_k = \{k_1, k_2 \cdots k_n\}$를 검색하며, 여기서 k_i는 U_d에서 i번째 문헌 k를 대출한 전체 대출자수가 된다.

(2) 협업 필터링을 이용한 도서 추천

협업 필터링을 이용한 추천기법에서 가장 필요한 정보는 이용자의 선호도 정보이다. 대출관련 정보 중에 대출일수는 한 문헌에 대한 이용자의 암묵적 선호도가 될 수 있다. 따라서 이러한 문헌에 대한 이용자의 선호도를 그 문헌의 대출일수로 대체하고, 대출일수를 이용한 협업 필터링 추천 시스템을 구축하였다.

이 연구에서는 희박함의 문제를 해결하기 위해‘ 아이템 중심 협업 필터링’ 기법을 적용하였다. 대출일수를 이용한 이 기법의 추천 과정은 다음과 같다. 먼저 이용자-문헌 행렬을 작성한다. 즉 문헌 d를 $U_d = \{u_{1d}, u_{2d} \cdots u_{nd}\}$와 같이 벡터 형태로 표현하면 u_{id}는 i번째 이용자가 대출한 문헌 d의 대출일수를 나타낸다. 이 행렬로부터 질의 문헌과 다른 문헌간의 유사도를 계산하여 유사계수 값이 기준치를 넘는 문헌들을 대상으로 순위화된 추천 리스트를 제공한다.

질의 문헌과 가장 유사한 문헌을 찾기 위해 사용된 유사계수로는 협업 필터링 기법에서 주로 사용되는 피어슨 유사계수(Pearson's correlation coefficient)와 코사인 유사계수(cosine coefficient)를 이용하였다(Sarwar et al. 2001).

두 문헌간의 유사도를 측정하기 위해 아래와 같이 피어슨 유사계수 공식을 사용할 경우, i, j는 이용자 u가 공통으로 대출한 두 문헌을 나타내며, R_{ui}와 R_{uj}는 각각 문헌 i, j에 대해 이용자 u의 대출일수를 나타내고, $\bar{R}_i$와 $\bar{R}_j$는 각각 두 문헌 i, j의 평균 대출일수를 나타낸다.

$$SIM_pear(i,j) = \frac{\sum_{u \in U}(R_{ui}-\bar{R}_i)(R_{uj}-\bar{R}_j)}{\sqrt{\sum_{u \in U}(R_{ui}-\bar{R}_i)^2}\sqrt{\sum_{u \in U}(R_{uj}-\bar{R}_j)^2}}$$

또한 다음의 코사인 유사계수 공식에서 $\bar{R}_u$는 이용자 u의 평균 대출일수를 나타낸다.

$$SIM_cos(i,j) = \frac{\sum_{u \in U}(R_{ui}-\bar{R}_u)(R_{uj}-\bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{ui}-\bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{uj}-\bar{R}_u)^2}}$$

(3) 내용기반 필터링을 이용한 도서 추천

내용기반 추천 기법은 문헌집단 내에서 질의 문헌과 다른 문헌간의 유사도를 측정하여 가장 유사도가 높은 문헌을 추천한다. 문헌간의 유사도 측정은 유사계수 공식을 사용하며 가장 기본적인 유사계수 공식으로 코사인 유사계수를 들 수 있다.

다음의 코사인 유사계수 공식에서 w_{ij}는 문헌 d_i에서 j번째 용어의 가중치를 나타내며, w_{iq}는 질의 문헌 q에서 j번째 용어의 가중치를 나타낸다(Baeza-Yates, and Ribeiro-Neto 1999).

$$sim(d_i, q) = \frac{\sum_{j=1}^{t} w_{ij} \times w_{iq}}{\sqrt{\sum_{j=1}^{t} w_{ij}^2} \times \sqrt{\sum_{j=1}^{t} w_{iq}^2}}$$

또한 용어 가중치로는 단어빈도 $w_{ij}(1)$과 가장 일반적으로 널리 쓰이는 상대빈도 개념인 역문헌빈도 $w_{ij}(2)$를 사용하였다. 아래의 식에서 f_{ij}는 문헌 d_i에서 j번째 용어의 출현횟수, N은 전체문헌의 수, n_j는 j번째 용어의 문헌빈도를 나타낸다.

$$w_{ij}(1) = f_{ij}$$

$$w_{ij}(2) = f_{ij} \times \log \frac{N}{n_j}$$

4. 실험 결과 분석

4.1 연관성 규칙을 이용한 추천 실험 결과

연관성 규칙을 이용한 추천 기법의 실험은 단순 대출빈도를 이용한 연관성 추천 기법(연관성_LF)과 대출자수를 이용한 연관성 추천 기법(연관성_UF)으로 나누어 추천 성능을 평가하였다. 각 방법에 따른 실험의 결과는 <표 2>와 같다. 연관성_LF의 평균재현율은 0.2265, 평균정확률은

0.1833으로 나타났으며, 연관성_UF의 평균 재현율은 0.4786, 평균정확률은 0.4200으로 나타났다.

이 실험 결과 연관성 규칙을 이용한 추천은 질의 문헌과 추천대상 문헌간의 연관성 측정을 위해 무슨 값을 사용했는가에 의해 큰 영향을 받는 것으로 나타났다. 특히 대출빈도를 이용한 연관성 추천 기법이 대출자수를 이용한 기법보다 매우 낮은 성능을 보이고 있는 이유 중의 하나는 문헌을 순위화 할 때 한 명의 이용자에 의해 여러 번 대출된 문헌은 그 만큼 높은 가중치를 얻어 높은 문헌 순위를 받기 때문인 것으로 보인다.

각 질의별 분석에 의하면 두 기법에서 모두 높은 재현율을 보인 14, 18, 27번 질의 문헌은 이들 문헌의 서명에 " 의료보장"," 행동수정"," 생활풍수"와 같은 비교적 특정성이 높은 용어가 포함되어 있었다.

<표 2> 연관성 규칙 기반 기법의 성능 평가

질문번호	이용자수	대출빈도	연관성_LF		연관성_UF	
			재현율	정확률	재현율	정확률
1	17	23	0.1429	0.2000	0.3571	0.5000
2	16	26	0.1000	0.1000	0.5000	0.5000
3	15	25	0.1667	0.1000	0.3333	0.2000
4	14	20	0.1000	0.1000	0.4000	0.4000
5	13	17	0.1429	0.1000	0.7143	0.5000
6	13	17	0.0833	0.1000	0.4167	0.5000
7	12	18	0.2222	0.2000	0.6667	0.6000
8	12	17	0.1667	0.3000	0.3333	0.6000
9	11	12	0.1667	0.1000	0.6667	0.4000
10	11	17	0.0000	0.0000	0.5000	0.3000
11	10	14	0.1667	0.1000	0.6667	0.4000
12	10	15	0.0625	0.1000	0.5000	0.8000
13	10	11	0.3000	0.3000	0.7000	0.7000
14	9	14	0.4000	0.2000	0.8000	0.4000
15	9	9	0.3333	0.2000	0.6667	0.4000
16	9	20	0.0526	0.1000	0.5263	1.0000
17	9	9	0.2222	0.4000	0.3889	0.7000
18	8	10	0.5000	0.2000	0.7500	0.3000
19	8	19	0.1818	0.2000	0.1818	0.2000
20	8	12	0.1250	0.1000	0.1250	0.1000
21	8	8	0.2500	0.2000	0.1250	0.1000
22	7	12	0.1333	0.2000	0.3333	0.5000
23	7	9	0.3750	0.3000	0.3750	0.3000
24	7	9	0.6000	0.3000	0.6000	0.3000
25	7	9	0.1429	0.1000	0.5714	0.4000
26	6	10	0.2308	0.3000	0.2308	0.3000
27	6	10	0.8333	0.5000	0.8333	0.5000
28	6	9	0.1667	0.1000	0.6667	0.4000
29	6	12	0.1429	0.1000	0.1429	0.1000
30	6	10	0.2857	0.2000	0.2857	0.2000
평균	9.7	14.1	0.2265	0.1833	0.4786	0.4200

<표 3> 협업 필터링 기법의 성능 평가

질문번호	이용자수	협업_pear		협업_cos	
		재현율	정확률	재현율	정확률
1	17	0.2857	0.4000	0.3571	0.5000
2	16	0.4000	0.4000	0.3000	0.3000
3	15	0.3333	0.2500	0.3333	0.2500
4	14	0.7000	0.7000	0.7000	0.7000
5	13	0.5714	0.4000	0.5714	0.4000
6	13	0.4167	0.5000	0.4167	0.5000
7	12	0.5556	0.5000	0.4444	0.4000
8	12	0.5000	0.9000	0.4444	0.8000
9	11	0.6667	0.4000	0.5000	0.3000
10	11	0.3333	0.2000	0.5000	0.3000
11	10	0.6667	0.5000	0.6667	0.5000
12	10	0.4375	0.7000	0.5000	0.8000
13	10	0.7000	0.7778	0.8000	0.8889
14	9	0.8000	0.4000	1.0000	0.5000
15	9	0.6667	0.4000	0.6667	0.4000
16	9	0.4211	0.8000	0.5263	1.0000
17	9	0.3889	0.7000	0.3333	0.6000
18	8	0.7500	0.3750	0.7500	0.3750
19	8	0.4545	0.5000	0.4545	0.5000
20	8	0.2500	0.2000	0.2500	0.2000
21	8	0.1250	0.1000	0.2500	0.2000
22	7	0.2667	0.4000	0.2667	0.4000
23	7	0.3750	0.3750	0.3750	0.3750
24	7	0.6000	0.7500	0.6000	0.7500
25	7	0.4286	0.3000	0.4286	0.3000
26	6	0.5385	0.7000	0.4615	0.6000
27	6	0.8333	0.8333	0.8333	0.8333
28	6	0.6667	0.5000	0.6667	0.5000
29	6	0.1429	0.1667	0.1429	0.1667
30	6	0.5714	0.4000	0.5714	0.4000
평균		0.4949	0.4843	0.5037	0.4913

4.2 협업 필터링을 이용한 추천 실험 결과

협업 필터링을 이용한 추천 기법의 실험은 피어슨 유사계수를 이용한 협업 필터링 기법(협업_pear)과 코사인 유사계수를 이용한 협업 필터링 기법(협업_cos)으로 나누어 추천 성능을 평가하였다. <표 3과> 같이 협업_pear 기법의 평균재현율은 0.4949, 평균정확률은 0.4843이며, 협업_cos 기법의 평균재현율은 0.5037, 평균정확률은 0.4913으로 나타났다. 특히 질의별로 재현율을 살펴보면 앞의 연관성 규칙을 이용한 추천 기법에서와 마찬가지로 질의번호 14, 18, 27번에서 성능이 비교적 높게 나타난 것을 볼 수 있다.

<표 3>에서 비슷한 추천 성능을 보이는 두 가지 방법의 순위화 성능을 비교해 보면 코사인 유사계수를 이용한 협업 필터링 추천 기법이 더 나은 결과를 보여준다. 협업_pear 기법은 총 184건의 적합문헌을 추천하였으며, 협업_cos 기법은 총 189건의 적합 문헌을 추천하였다. 하지만 순위 3위 내에 드는 각 기법의 적합문헌 수는 협업_pear 기법이 65건, 협업_cos 기법이 77건으로 협업_cos 기법이 12건이나 더 많았다. 대부분의 이용자들은 검색 결과 중에 10건의 문헌만을 살펴보는 경우가 많다. 특히, 일부 이용자는 상위 2~3건의 문헌을 살펴보는 경향이 있으므로 순위화 성능이 높은 추천 기법일수록 이용자에게 도움이 될 것이다.

4.3 내용기반 필터링을 이용한 추천 실험 결과

각각 단어빈도(내용_tf)와 역문헌빈도(내용_idf) 가중치를 이용하여 문헌간 유사도를 측정한 내용기반 추천 기법의 성능 평가는 <표 4>와 같다. 전체적으로 내용기반 추천 기법은 다른 추천 기법들에 비해 월등히 좋은 결과를 보여주고 있다. 즉, 가중치 값으로 단어빈도를 이용한 실험에서는 평균재현율과 평균정확률이 각각 0.5201과 0.5010으로 나타났으며, 역문헌빈도를 이용한 실험에서는 평균재현율과 평균정확률이 0.5539와 0.5343으로 나타났다. 두 가지 가중치 방법을 비교하면 재현율과 정확률 모두에서 역문헌빈도를 이용한 방법(내용_idf)이 7%정도 높은 성능을 보이고 있다.

<표 4>를 보면 다른 질의보다 높은 성능을 보이는 질의는 6, 20, 21, 29번인데 이들은 모두 일반적인 개론서에 해당한다. 이런 결과는 내용기반 추천 기법이 키워드를 통해 유사한 문헌을 검색하기 때문에 특정 주제에 대한 개론서의 경우 그만큼 검색될 확률이 커지기 때문이다. 반면 <표 4>에서 추천 성능이 평균 성능보다 현저히 낮은 질의인 5, 14, 18번의 경우는 그 이유가 용어의 특정성 때문인 것으로 보인다. 이와 같이 내용기반 추천 기법의 성능은 전체적으로 볼 때 질의별 편차가 큰 것으로 나타나 있다.

4.4 혼합형 필터링을 이용한 추천 실험 결과

<표 4> 내용기반 필터링 기법의 성능 평가

질문 번호	이용 자수	내용_tf		내용_idf	
		재현율	정확률	재현율	정확률
1	17	0.5714	0.8000	0.6429	0.9000
2	16	0.6000	0.6000	0.6000	0.6000
3	15	0.5000	0.4286	0.5000	0.4286
4	14	0.5000	0.5000	0.5000	0.5000
5	13	0.1429	0.1000	0.1429	0.1000
6	13	0.7500	0.9000	0.8333	1.0000
7	12	0.3333	0.3000	0.3333	0.3000
8	12	0.5000	0.9000	0.5556	1.0000
9	11	1.0000	0.6000	0.8333	0.5000
10	11	0.6667	0.4000	0.6667	0.4000
11	10	0.5000	0.3000	0.8333	0.5000
12	10	0.5000	0.8000	0.5000	0.8000
13	10	0.9000	0.9000	0.9000	0.9000
14	9	0.0000	0.0000	0.2000	0.1000
15	9	0.3333	0.2000	0.5000	0.3000
16	9	0.4211	0.8000	0.4737	0.9000
17	9	0.5556	1.0000	0.5556	1.0000
18	8	0.5000	0.2000	0.2500	0.1000
19	8	0.4545	0.5000	0.5455	0.6000
20	8	0.8750	0.7000	0.7500	0.6000
21	8	0.7500	0.6000	0.7500	0.6000
22	7	0.5333	0.8000	0.6000	0.9000
23	7	0.2500	0.2000	0.3750	0.3000
24	7	0.6000	0.3000	0.6000	0.3000
25	7	0.7143	0.5000	0.7143	0.5000
26	6	0.4615	0.6000	0.4615	0.6000
27	6	0.5000	0.3000	0.5000	0.3000
28	6	0.3333	0.2000	0.5000	0.3000
29	6	0.7143	0.5000	0.7143	0.5000
30	6	0.1429	0.1000	0.2857	0.2000
평균		0.5201	0.5010	0.5539	0.5343

<표 5> 혼합형 필터링 기법의 성능 평가

질문 번호	이용 자수	혼합_mix		혼합_swtch	
		재현율	정확률	재현율	정확률
1	17	0.4286	0.6000	0.3571	0.5000
2	16	0.7000	0.7000	0.3000	0.3000
3	15	0.3333	0.2000	0.5000	0.4286
4	14	0.8000	0.8000	0.7000	0.7000
5	13	0.2857	0.2000	0.5714	0.4000
6	13	0.6667	0.8000	0.8333	1.0000
7	12	0.6667	0.6000	0.3333	0.3000
8	12	0.5556	1.0000	0.4444	0.8000
9	11	0.8333	0.5000	0.8333	0.5000
10	11	0.6667	0.4000	0.6667	0.4000
11	10	0.8333	0.5000	0.8333	0.5000
12	10	0.5625	0.9000	0.5000	0.8000
13	10	0.7000	0.7000	0.8000	0.8889
14	9	0.4000	0.2000	1.0000	0.5000
15	9	0.6667	0.4000	0.6667	0.4000
16	9	0.4737	0.9000	0.5263	1.0000
17	9	0.4444	0.8000	0.5556	1.0000
18	8	0.5000	0.2000	0.2500	0.1000
19	8	0.4545	0.5000	0.5455	0.6000
20	8	0.5000	0.4000	0.7500	0.6000
21	8	0.7500	0.6000	0.7500	0.6000
22	7	0.4667	0.7000	0.6000	0.9000
23	7	0.3750	0.3000	0.3750	0.3000
24	7	1.0000	0.5556	0.6000	0.3000
25	7	0.7143	0.5000	0.7143	0.5000
26	6	0.3077	0.4000	0.4615	0.6000
27	6	1.0000	0.6000	0.8333	0.8333
28	6	0.5000	0.3000	0.6667	0.5000
29	6	0.4286	0.3000	0.7143	0.5000
30	6	0.5714	0.4000	0.5714	0.4000
평균		0.5862	0.5319	0.6085	0.5717

혼합형 필터링 추천 기법은 하나 이상의 추천 기법을 조합하여 각 추천 기법의 문제점을 보완함으로써 보다 좋은 성능을 가져올 수 있다. 따라서 이 연구에서는 협업 필터링을 이용한 추천 기법과 내용기반 필터링을 이용한 추천 기법을 혼합하여 추천 성능을 증진시키는 방법을 제시하고자 한다.

우선 협업 필터링을 이용한 기법과 내용기반 필터링을 이용한 기법의 추천 결과 중에 각각 상위 5건씩을 혼합하여 최종 추천 결과로 제공하는 방법(혼합_mix)을 실험하였다. 다른 방법(혼합_swtch)에서는 내용기반 필터링을 이용한 추천 기법을 주로 적용하되, 질의 문헌의 서명에서 추출된 키워드 중에 하나라도 문헌빈도가 10이하일 경우 협업 필터링 추천 기법을 적용하였다.

혼합형 필터링을 이용한 기법의 추천 성능은 <표 5>와 같다. 혼합_mix 기법의 평균재현율은 0.5832이며 평균정확률은 0.5319로 나타났다. 평균정확률의 경우 내

용기반 필터링을 이용한 추천 기법의 평균 정확률인 0.5343보다 낮게 나타났다. 그러나 혼합_mix 기법에서 질의별 편차는 줄어들었다.

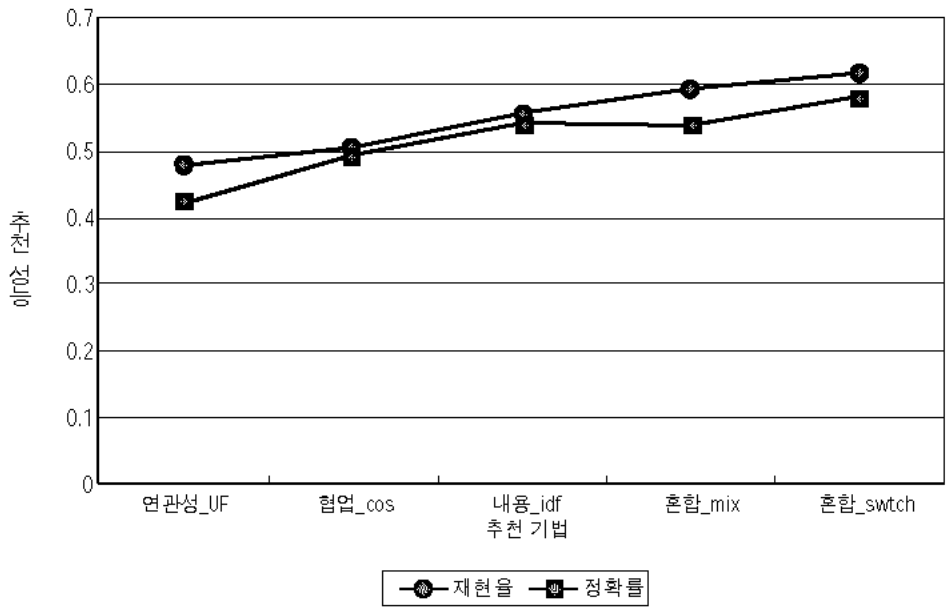
<표 5>를 보면 혼합_swtch 기법의 평균 재현율은 0.6085이며 평균정확률은 0.5717로 나타났다. 이 기법의 경우 다른 모든 기법들보다 더 좋은 성능을 보임을 알 수 있다. 따라서 내용기반 필터링 추천 기법의 성능이 심각히 저하되는 질의 문헌의 경우 협업 필터링을 이용한 추천 기법을 적용하거나, 한번도 대출되지 않은 도서처럼 협업 필터링 추천 기법을 이용하기 어려울 경우 내용기반 필터링 추천 기법을 적용하는 혼합형 필터링 추천 기법을 통해 보다 좋은 추천 성능을 얻을 수 있음을 알 수 있다.

4.5 실험결과의 요약 및 비교 분석

앞에서 평가한 세 가지 추천 기법과 혼합형 추천 기법의 성능을 비교하면 <표 6> 및 <그림 1>과 같다.

<표 6> 추천 기법의 성능 비교

	연관성_UF	협업_cos	내용_idf	혼합_mix	혼합_swtch
재 현 율	0.4786	0.5037	0.5539	0.5862	0.6085
정 확 률	0.4200	0.4913	0.5343	0.5319	0.5717



<그림 1> 추천 기법에 따른 성능 비교

먼저 연관성_UF 기법은 다른 기법에 비해 비교적 낮은 성능을 보였다. 그러나 추천 성능은 연관성 측정 방법에 의해 큰 영향을 받으므로 보다 정확하게 연관성 정도를 산출할 수 있다면 이 기법의 추천 성능은 충분히 향상될 가능성이 있다. 이 추천 기법은 다른 기법에 비해 간단하게 구축할 수 있으므로 실제 도서관 시스템에서 이용자에게 실시간으로 추천 서비스를 제공하기가 용이하며, 비교적 적은 자원으로 의미 있는 결과를 얻을 수 있다는 장점이 있다. 도서 추천의 정확성을 높일 수 있는 방안으로 비교적 오랜 기간 동안 대출정보를 축적할 수 있다면 보다 좋은 추천 성능을 가져올 수 있을 것이다.

협업 필터링 추천 기법은 추천할 대상의 표현에 의존할 필요가 없다. 즉 질의 문헌의 키워드와 동등관계에 있는 동의어나 유사동의어를 이용해 질의를 확장 검색하는 것과 비슷한 결과를 가져온다. 예를 들어 질의번호 22번의 경우 주요 키워드인 " information"과" retrieval" 이외에도 한글로 된 키워드인" 정보검색"이 포함된 문헌이 검색된다. 또한" 실험설계"라는 키워드를 갖는 문헌을 질의로 입력하면" 신뢰도와 타당도"라는 서명의 문헌이 검색된다. 이러한 분석 결과를 본다면 실제 구축된 도서 추천 시스템에서 협업 필터링 기법은 이 실험에서보다 더 좋은 성능을 보여줄 것으로 예상된다. 왜냐하면 이 기법은 질의어와 정확히 일치하는 문헌만 검색하는 것이 아니고 이용자의 정보요구와 관련된 폭넓고 다양한 문헌들을 추천하는 특성을 갖고 있기 때문이다.

협업 필터링 추천 기법의 문제점은 문헌의 대출자수가 적을 경우 추천 성능이 떨어지거나 아예 추천이 불가능해진다는 것이다. 따라서 대출이 적게 되거나 몇몇 소수의 이용자에게 집중되어 대출되는 문헌의 경우 추천이 어려워진다. 더 나아가 전체 장서 중에 극히 소수의 문헌만 집중적으로 이용되는 경우 이 기법을 단독적으로 적용하는 것은 바람직하지 않다.

내용기반 필터링 추천 기법은 다른 단일 추천 기법보다 우수한 성능을 나타냈다. 내용기반 필터링 기법의 평균재현율은 0.5549로서 연관성 규칙을 이용한 기법보다 15.73%, 협업 필터링 기법보다 9.97%의 성능 향상을 보이고 있으며, 또한 평균정확률은 0.5343으로서 연관성 규칙 기반 기법보다 27.21%, 협업 필터링 기법보다 8.75%가 향상되었다.

내용기반 추천 기법은 아이템의 내용에 기초하여 추천하기 때문에 이용자나 다른 요인의 영향을 받지 않는다. 특히 아이템이 문헌일 경우 내용 분석이 필수적으로 이루어져야 하기 때문에 내용기반 필터링 추천 기법이 가장 적용하기 쉽고 보편적인 추천 기법 중의 하나라고 할 수 있다. 그러나 내용기반 추천 기법의 경우 추천하기 위해 이용된 주요 키워드가 낮은 출현빈도를 갖거나 너무 일반적인 성격의 키워드가 질의에 많이 포함된다면 추천 성능이 떨어질 수밖에 없다.

혼합형 필터링을 이용한 추천 기법의 성능은 단일 추천 기법보다 높게 나타났다. 또한 <표 6>에서와 같이 혼합형 필터링 추천 기법의 두 가지 방법 중에서는 혼합

_swtch 방법이 더 우수한 성능을 보였다. 연관성 규칙 기반 기법, 협업 필터링 기법, 내용기반 필터링 기법에 비해 혼합_swtch 기법은 평균재현율은 각각 27.14%, 20.81%, 9.86% 향상되었고 평균정확률은 36.12%, 16.36%, 7.00% 향상되었다.

실험 결과 대출대상 도서의 추천에 있어 혼합형 필터링 기법이 협업 필터링 기법이나 내용기반 필터링 기법을 단독적으로 적용하는 것보다 시간적인 면에서는 다소 비효율적이지만 성능면에서는 매우 우수한 것으로 나타났다.

5. 결 론

이 연구에서는 도서관에 소장된 정보를 이용자들에게 보다 적극적으로 제공할 목적으로 기존의 도서관 시스템에 추천 기법을 도입하여 이용자 관심 분야의 도서를 자동으로 추천하는 시스템을 구현하였다. 도서 추천 시스템에서 가장 좋은 성능을 보이는 추천 기법을 찾아내기 위해 연관성 규칙 기반 기법, 협업 필터링 기법, 내용기반 필터링 기법, 그리고 혼합형 기법을 사용한 실험을 수행하였다.

도서 추천 실험 결과 세 가지 단일 기법 가운데 문헌간 연관성 규칙을 이용한 추천 기법의 성능이 가장 낮았고 내용기반 필터링 기법이 가장 좋은 성능을 보였다.

협업 필터링을 이용한 추천 기법의 경우 피어슨 유사계수와 코사인 유사계수를 이용하여 두 문헌간의 유사도를 측정한 결과 코사인 유사계수를 이용한 방법이 보다 좋은 성능을 나타냈으며, 순위화 성능에서도 더 좋은 결과를 보였다. 특히 협업 필터링을 이용한 추천 기법은 추천할 대상의 표현에 의존하지 않기 때문에 입력된 질의 문헌에서 추출된 키워드와 동등관계에 있는 동의어나 유사동의어를 갖는 문헌들이 함께 추천되는 특징을 보여주었다.

내용기반 추천 기법의 경우 코사인 유사계수를 이용하였으며 용어 가중치 기법으로 단어빈도와 역문헌빈도를 이용하였다. 그 결과 다른 단순 기법 보다 우수한 추천 성능을 보였으며 역문헌빈도를 이용한 기법이 단어빈도를 이용한 경우보다 더 높은 성능을 보였다.

마지막으로 혼합형 필터링 추천 기법은 다른 단일 추천 기법 보다 추천 성능이 높게 나타났다. 특히 문헌빈도에 따라 다른 단일 추천 기법을 적용한 방법이 더 우수한 성능을 보였으며, 각각 연관성 규칙 기반 기법, 협업 필터링 기법, 내용기반 필터링 기법에 비해 평균재현율에 있어 27.14%, 20.81%, 9.86% 향상되었고, 평균정확률은 36.12%, 16.36%, 7.00% 향상되었다.

결론적으로 도서관 시스템의 대출정보와 서지정보를 이용하여 대출대상 도서를 추천하는 데 있어 협업 필터링 추천 기법과 내용기반 필터링 추천 기법을 각각 따로 적용하는 것보다 두 기법을 함께 이용한 혼합형 필터링 추천 기법이 가장 효과적임을 알 수 있다.

참고문헌

Baeza-Yates, Ricardo, and Berthier Ribeiro-Neto. 1999. *Modern Information* Retrieval. New York : Addison-Wesley.

Balabanovic, Marko, and Yoav Shoham. 1997." Fab: Content-based, Collaborative Recommendation". *Communications of the ACM,* 40(3) : 66-72.

Breese, John S., and David Heckerman, Carl Kadie. 1998. " Empirical Analysis of Predictive Algorithms for Collaborative Filtering". *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence.*

Burke, Robin. 2001." Hybrid Recommender Systems: Survey and Experiments". *User Modeling and User-Adapted Interaction.* [online]. [cited 2001. 12.09]. <http://cbeblade.fullerton.edu/~rburke/pubs/burke-umuai-ip.pdf>.

Funakoshi, Kaname, and Takeshi Ohguro. 2000." A Content-based Collaborative Recommender System with Detailed Use of Evaluations". *Fourth International Conference on Knowledge-based Intelligent Engineering Systems & Allied Technologies,* 253-256.

Goldberg, David, David Nichols, Brian M. Oki, and Douglas Terry. 1992. " Using Collaborative Filtering to Weave an Information TAPESTRY". *Communcation of the ACM,* 35(12) : 61-70.

Good, Nathaniel, Ben Schafer, Joseph A. Konstan, Al Borchers, Badrul Sarwar, Jon Herkocker, and John Riedl. " Combining Collaborative Filtering with Personal Agents for Better Recommendations". *Proceedings of AAAI-99,* 439-446.

Herlocker, Jonathan L., Joseph A. Konstan, Al Borchers, and John Riedl. 1999. " An Algorithmic Framework for Performing Collaborative Filtering". *Proceedings of the 1999 Conference on Research and Development in Information Retrieval,* 230-237.

Hill, Will, Larry Stead, Mark Rosenstein, and George Furnas. 1995." Recommending and Evaluating Choices in a Virtual Community of Use". *Proceedings of CHI'95,* 194-201.

Melville, Prem, Raymond J. Mooney, and Ramadass Nagarajan. 2001." Content-Boosted Collaborative Filtering". *Proceedings of the SIGIR-2001 Workshop on Recommender Systems.* [online]. [cited 2001.12.03]. <http://cs.oregonstate.edu/~herlock/rsw2001/final/full_length_papers/2_Melville+.pdf>.

Resnick, Paul, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John

Riedl. 1994.“ GroupLens: An Open Architecture for Collaborative Filtering of Netnews”. *Proceedings of the 1994 Computer Supported Cooperative Work Conference,* 175-186

Sarwar, Badrul M., George Karypis, Joseph A. Konstan, and John T. Riedl. 2000.“ Application of Dimensionality Reduction in Recommender System - A Case Study”. *Proceedings of WebKDD 2000 Workshop Web Mining for E-Commerce- Challenges and Opportunities.* [online]. [cited 2001.11.09]. <http://robotics.stanford.edu/~ronnyk/WEBKDD2000/papers/sarwar.pdf>.

Sarwar, Badrul, George Karypis, Joseph Konstan, and John Riedl. 2001.“ Item-based Collaborative Filtering Recommendation Algorithms”. *Proceedings of the 10th International World Wide Web Conference,* 285 - 295.

Schafer, J. Ben, Joseph A. Konstan, and John Riedl. 1999.“ Recommender System in E-commerce”. *Proceedings of the First ACM Conference on Electronic Commerce,* 158-166.

Shardanand, Upendra, and Pattie Maes. 1995“ Social Information Filtering: Algorithms for Automating‘ Word of Mouth’”. *Proceedings of the CHI'95,* 210-217.

Terveen, Loren, Will Hill, Brian Amento, David McDonald, and Josh Creter. 1997.“ PHOAKS: A System for Sharing Recommendations”. *Communciation of the ACM,* 40(3) : 59-62.