정보관리연구, vol.41, no.4 2010, pp.227-249

나이브베이즈 분류모델과 협업필터링 기반 지능형 학술논문 추천시스템 연구

A Study of Intelligent Recommendation System based on Naive Bayes Text Classification and Collaborative Filtering

이 상 기* · 이 병 섭** · 박 병 용*** · 황 혜 경****
Sang-Gi Lee · Byeong-Seop Lee · Byeong-Yong Bak · Hye-Kyong Hwang

차 례

1. 서론
2. 연구배경
3. 관련연구
4. 제안모델
5. 실험결과
6. 결론
· 참고문헌

* 한국과학기술정보연구원 정보서비스실 선임연구원
(Senior Researcher, Dept. of Information Service, KISTI, sklee@kisti.re.kr)
** (주)아르고넷 개발팀장
(Development Team Leader, ARGONET, bslee73@gmail.com)
*** 성균관대학교 문헌정보학과 석사과정, KISTI 정보서비스실 인턴연구원
(Graduate Student, Dept. of Library and Information Science, Sungkyunkwan University and Research, Internship, KISTI, byoungyong.baag@kisti.re.kr)
**** 한국과학기술정보연구원 지식기반실 선임연구원
(Senior Researcher, Dept. of Knowledge Resources, KISTI, hkhwang@kisti.re.kr)
• 논문접수일자: 2010년 6월 25일
• 최종심사일자: 2010년 9월 30일
• 게재확정일자: 2010년 10월 11일

초 록

정보기술과 인터넷의 발달로 학술정보가 폭발적으로 증가하고 있다. 정보 과잉으로 인해 연구자들은 필요한 정보를 찾거나 필터링하는데 더 많은 시간과 노력을 투입하고 있다. 이용자들이 원하는 정보를 예측하여 관심 가질만한 정보를 선별하여 추천하는 시스템을 전문가시스템, 데이터마이닝, 정보검색 등 다양한 분야에서 오래 전부터 연구하여 왔다. 최근에는 콘텐츠기반추천시스템과 협업필터링을 결합하거나 다른 분야 모델을 접목한 하이브리드 추천시스템으로 발전하고 있다. 본 연구에서는 기존 추천시스템 문제를 해결하고 대규모 정보센터나 도서관에서 학술논문을 효율적이고 지능적으로 추천하기 위해 협업필터링과 나이브베이즈모델을 결합한 새로운 방식의 추천시스템을 제시하였다. 즉, 협업필터링 방식으로 과도한 특성화(Over-specialization) 문제를 해결하고, 나이브베이즈모델을 통해 평가정보나 이용정보가 부족한 신규콘텐츠 추천문제를 해소하였다. 본 모델을 검증하기 위해 한국과학기술정보연구원 NDSL에서 제공하는 식품과 전기 분야 학술논문에 적용하여 실험하였다. 현재 NDSL 이용자 4명에게 피드백을 받은 결과 추천논문에 상당히 만족하는 것으로 나타났다.

키 워 드

추천시스템, 하이브리드추천시스템, 나이브베이즈모델, 협업필터링

ABSTRACT

Scholarly information has increased tremendously according to the development of IT, especially the Internet. However, simultaneously, people have to spend more time and exert more effort because of information overload. There have been many research efforts in the field of expert systems, data mining, and information retrieval, concerning a system that recommends user-expected information items through presumption. Recently, the hybrid system combining a content-based recommendation system and collaborative filtering or combining recommendation systems in other domains has been developed. In this paper we resolved the problem of the current recommendation system and suggested a new system combining collaborative filtering and Naive Bayes Classification. In this way, we resolved the over-specialization problem through collaborative filtering and lack of assessment information or recommendation of new contents through Naive Bayes Classification. For verification, we applied the new model in NDSL's paper service of KISTI, especially papers from journals about Sitology and Electronics, and witnessed high satisfaction from 4 experimental participants.

KEYWORDS

Recommendation System, Hybrid Recommendation System, Naive Bayes Classifiers, Collaborative Filtering

1. 서론

다양한 저작도구(authoring tool)의 발달과 인터넷 확산에 따른 정보과잉(information overflow)으로 인해 연구자들은 원하는 정보를 찾기 위해 점점 더 많은 시간을 소비하고 노력을 기울이고 있다. 특히 대규모 학술정보를 서비스하는 정보센터나 도서관에서는 이런 현상이 더욱 심하다. 정보검색 분야에서는 오래전부터 키워드 출현빈도에 기반 한 적합문서 선별 및 랭킹, 이용로그 및 이용 상황에 근간한 정보 필터링, 검색결과 클러스터링 및 분석, 페이지 랭크 등 다양한 연구를 통해 이런 문제에 대응해 왔다. 정보센터나 도서관에서도 검색 질문식이나 즐겨 찾는 저널을 등록한 후 새로운 정보가 입수되면 자동으로 관련 정보를 제공하는 SDI(Selective Dissemination of Information) 서비스를 제공하고 있다. 하지만 SDI의 경우 등록 및 갱신절차가 번거롭고 정보량에 비례하여 불필요한 정보가 많이 포함됨으로써 만족도가 떨어지고 있다. 연구자들은 연구 시 정보를 찾는데 많은 시간을 투입하고 있으며, 연구자들이 관심 가질 만한 학술논문을 지능적으로 선별하여 추천하면, 연구자들이 정보를 찾는데 드는 노력을 절감하고 만족도도 제고할 수 있다.

최근 고객들이 원하는 정보를 필터링하여 제공하는 개인화서비스와 이용자들이 선호할 만한 정보를 예측하여 제공하는 추천시스템에 대한 연구가 활발하다. 콘텐츠 부문의 추천시스템은 이용자 프로파일, 이용로그, 관심사항, 콘텐츠 유사도 등을 토대로 적합문서를 지능적으로 추천하는 시스템이다. 추천방식으로는 콘텐츠 유사도에 근간한 콘텐츠기반추천시스템과 같은 분야를 연구하는 동료 연구자들이 이용한 정보를 토대로 추천하는 협업필터링이 있으며, 최근 이들을 결합한 새로운 방식의 하이브리드 추천모델이 등장하고 있다.

본 연구에서는 기존 추천시스템 문제점을 해결하고 대규모 정보센터나 도서관에서 학술논문을 효율적, 지능적으로 추천하기 위해 협업필터링과 나이브베이즈모델을 결합한 새로운 방식의 추천시스템을 제시하였다. 기존 추천시스템들의 경우 평가정보나 이용정보 부족, 과도한 특성화(Over-specialization) 문제, 비효율성 등 여러 가지 문제가 있다. 본 연구에서는 협업필터링 방식을 통해 과도한 특성화 문제를 개선하였으며, 나이브베이즈모델을 결합하여 이용로그 등이 축적될 때까지 기다려야하는 문제를 해결하였다. 특히, 협업필터링을 통해 문서학습 및 자질선정을 자동화함으로써 연구자들이 별도로 프로파일을 갱신하거나 관리할 필요가 없도록 하였다.

본 모델을 검증하기 위해 한국과학기술정보연구원 NDSL 학술논문을 대상으로 실험모형을 개발한 후 추천품질을 검증하였다. 실험대상은 최근 3년간 NDSL 이용로그를 토대로 이용 빈도가 높은 학술논문 약 3천 5백 건을 추출하고 이를 많이 이용한 식품과 전기 분야 연구자 상위 5명을 선정하여 추천논문에 대한

품질을 조사하였다. 이들에게 본 모델을 통해 추천한 논문을 이메일로 발송한 후 5점 척도로 만족도를 조사한 결과 대체로 만족하는 것으로 나타났다.

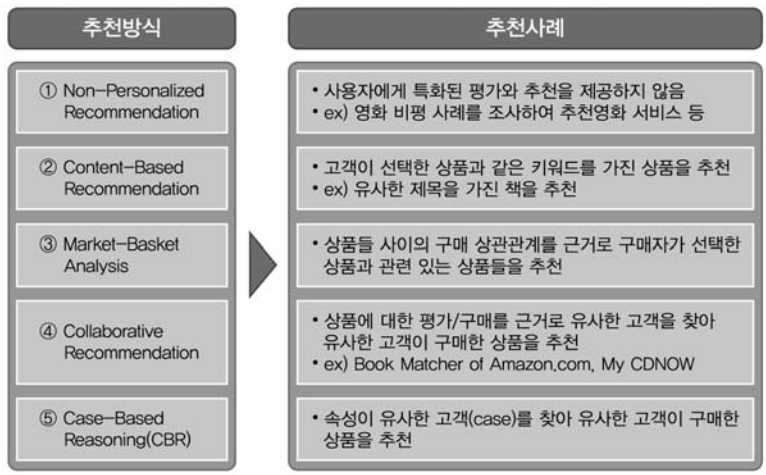
2. 연구배경

학술논문 등을 지능적으로 추천하기 위해 콘텐츠기반추천시스템(Content-Based Recommendation)과 협업필터링(Collaborative Filtering) 방식이 동시에 발전하여 왔으나, 현재는 후자에 관한 연구에 집중하고 있다. 추천시스템은 어떤 이용자와 특정 아이템이 주어졌을 때, 이 아이템에 대한 이용자의 선호도(preference)를 예측(prediction)하여 예상 선호도가 높은 아이템을 추천(recommendation)하거나(안신현 2007), 정보를 필터링(Information Filtering)하여 이용자가 흥미를 가질 만한 정보 아이템을 제공하는 시스템이다(한성희 2009).

추천방식은 〈그림 1〉과 같이 다양하며 대부분 콘텐츠나 사람 간 유사도에 기반하여 추천한다. 대표적인 추천시스템은 Amazon.com의 북매처(BookMatcher)다. 아마존닷컴에서는 개인별 특성이나 성향이 비슷한 이용자 집단을 분석하여 추천하는 방법 대신 상품 대 상품(Item-to-Item)에 기반한 추천방식을 사용하고 있다.

추천프로세스는 〈그림 2〉와 같이 정보를 수집하여 다양한 모델을 이용해 연관성을 분석하고 가장 적합한 문서를 계산 및 선별하여 추천한다. 추천정보는 검색시스템, 개인맞춤서비스, 타겟광고, 시장예측 등 다양한 분야에 활용된다.

이용자들이 원하는 정보를 예측하여 관심



〈그림 1〉 다양한 추천방식

〈그림 2〉 추천 프로세스(한성희 2009)

가질만한 정보를 선별하여 추천하는 시스템을 전문가시스템, 데이터마이닝, 정보검색 등 여러 분야에서 오래전부터 연구해 왔다. 특히, 정보검색의 경우 이용자 질의에 가장 적합한 문서를 상위에 추천하기 위해 키워드 출현빈도에 따라 가중치를 부여하거나 개인프로파일이나 이용로그를 토대로 검색결과를 필터링하고 있다.

최근 연구가 활발한 협업필터링 방식의 경우 성향이 비슷한 이용자들이 선호하는 정보를 추천하는 시스템으로 과도한 특성화(Over-specialization) 문제도 없고 추천 성능도 우수하다. 하지만 대다수 학술논문 서비스 사이트의 경우 콘텐츠를 평가하는 기능이 없으며 설사 기능이 있더라도 이용자들이 평가에 소극적이기 때문에 정보 희소성 문제에 봉착하게 된다. 게다가 신규 콘텐츠의 경우 일정한 시간이 경과되기 전까진 평가정보와 이용내역이 없어 추천할 수 없는 맹점이 있다. 이는 학술논문과 같이 시급성과 적시성이 중요한 부문에서는 이슈가 된다. 한편 콘텐츠기반추천 시스템은 문서 유사도에 기반하기 때문에 과도한 특성화 문제가 발생하며, 시간에 따라 바뀌는 이용자들의 관심사항을 적시에 반영할 수 없는 한계가 있다.

본 연구에서는 이와 같은 문제를 해결하고 대규모 정보센터나 도서관에서 지능적으로 학술논문을 추천할 수 있도록 상호 보완적인 모델을 결합하여 이를 해결하였다. 즉, 과도한 특성화 문제와 시시각각 바뀌는 이용자들의 관심사항을 자동으로 갱신할 수 있는 협업필터링모델에 신규 콘텐츠를 효율적으로 추천할 수 있도록 나이브베이즈모델을 접목하였다.

문서분류 방식에는 나이브베이즈모델 뿐만 아니라 다양한 모델이 있지만 Support Vector Machine(SVM)의 경우 두 개의 카테고리로 문서를 분류하는데 최적인 모델이므로 학술논문 추천시스템과 같이 다수의 이용자에게 분류하는 본 연구에는 적합하지 않다고 판단되어 제외하였으며, K-Nearest Neighbor(K-NN)의 경우 K 변수 선정문제로 고려하지 않았다.

3. 관련연구

3.1 추천시스템

추천시스템은 고객이 관심을 가지는 상품에 관한 정보나 인구통계학적 정보, 과거 구매한 행동분석을 토대로 고객의 요구에 맞는 상품을 추천해 주는 시스템이다(Sarwar 2001). 또한 고객들이 구매하고자 하는 상품을 쉽게 찾을 수 있도록 도와주는 정보필터링 기술이다(Schafer 1999). 추천시스템은 인터넷을 기반으로 고객별 일대일 마케팅을 가능하게 하는 e-CRM의 한 분야로서 Amazon, CD-now 등 해외 유수 전자상거래 사이트에서 활용되고 있으며, Ringo 음악추천이나 Bellcore 비디오 추천 등에 이용되고 있다.

Sarwar(2001)는 데이터 희박성(sparsity)과 규모의 문제를 지닌 이용자기반 협업필터링을 보완하기 위해 아이템기반 협업필터링 추천알고리즘을 제시하였다. 또한, 박지선(2002)은 아이템기반 협업필터링의 문제점을 보완하고자 유사 고객군을 찾은 다음 그들이 평가한 상품 간 유사도를 계산해서 추천하는 2-way 협업필터링을 제시하였다. 아울러 Kim(2002)은 K-means 클러스터링 기법을 추천알고리즘에 적용하였고, Li and Kim(2003)은 클러스터링기법을 아이템기반 협업필터링에 응용하였다. 이외에도 Roh(2003)의 연구에서는 클러스터링을 위하여 SOM을 사용하고 유사도를 찾기 위하여 최근접이웃법을 이용하여 기존의 협업필터링방식과 비교분석하여 예측 성능의 우수함을 설명하였다. 아울러 김재경(2003)은 전자상거래 쇼핑몰에 웹데이터마이닝과 클러스터링에 기반한 협업필터링 방안을 제시하였다. Weng(2004)은 고객이 새로운 제품을 아직 구매하지 않았기 때문에 기존의 장바구니 분석과 협업필터링 분석을 통해서는 새로운 제품을 추천할 수 없다고 보고 제품의 특징에 기반한 추천을 제안하였다. 끝으로 이충무(2009)는 같은 논문을 열람한 이용자를 유사한 분야를 연구하는 동료연구자로 간주하고 이들이 이용한 논문을 상호추천하고 정보를 공유하는 온라인 지식네트워크인 연구자연결망을 연구하였다.

가. 콘텐츠기반추천시스템

콘텐츠기반추천시스템(Content-Based Recommendation)은 정보검색기술에 바탕을 둔 시스템으로, 아이템의 콘텐츠를 직접 분석하여 아이템과 아이템, 아이템과 이용자 선호도 간 유사성을 분석하여, 이를 토대로 새로운 아이템을 추천해주는 시스템이다(안신현 2007). 본 방식은 분석의 용이함 때문에 주로 텍스트 기반의 뉴스나 인터넷 기사, 책 등을 추천하는 시스템에서 많이 사용하였으며(Krulwich 1995), 아이템의 메타데이터를 통해서도 내용을 분석할 수 있기 때문에 영화, 음악, 도서 등 다양한 문화 콘텐츠 추천시스템에서도 사용했다. 이용자 선호도에 관계없이 특정 아이템에 대한 유사 아이템 리스트를 제공하는 서비스가

가능한데, 음악 사이트에서 제공하는 유사 앨범이나 유사 아티스트가 여기에 해당한다.

Baumann(2005)은 2005년 아티스트에 관한 리뷰를 웹검색을 통해 수집하여 벡터 간 유사성 계산법으로 유사 아티스트를 찾아내는 연구를 하였다. 하지만 이용자의 취향은 개개인에 따라 다르기 때문에 특정 이용자에게 개인화된(Personalized) 추천리스트를 제공하기 위해, 이용자 선호도 프로파일(User Preference Profile)을 사용하였다. 즉, 이용자의 선호도를 각 아이템의 내용과 비교하여 선호도가 높을 것으로 예상되는 아이템을 추천하는 것이다. 이용자 선호도 프로파일은 주로 이용자가 과거에 이용했던 아이템에서 추출한다. 예를 들어 과거에 구매했던 도서나 영화, 음악, 읽었던 기사 또는 방문했던 URL 등이 이용자 프로파일이 되기도 하고, 직접적으로 특정 아이템에 대한 이용자 평가를 물어보기도 한다.

콘텐츠기반추천시스템은 이해하기 쉽고 계산과정이 간단한 반면 여러 가지 단점이 있다. 먼저, 다룰 수 있는 분야가 대부분 텍스트로 그 범위가 매우 한정적이다. 즉, 영화, 음악, 레스토랑과 같은 아이템에 포함된 영상, 음향, 맛 등은 적용하기 어렵다. 인터넷 텍스트 문서도 내용 파악은 가능하나, 웹페이지의 미적 감각이나 멀티미디어 정보, 로딩시간 등은 파악하기 어렵다. 둘째, 추천 아이템이 지나치게 특성화되기 싶다. 이용자 선호도와 유사한 아이템만 찾다 보면, 과거에 접했던 아이템과 다른 새로운 아이템을 접할 기회가 없게 된다(Balabanovic 1997). 이러한 문제를 해결하기 위해 추천에서 무작위 요소를 추가하거나 돌연변이 방식을 사용하기도 한다(Sheth, B. 1993). 셋째, 이용자 선호도 프로파일이 있어야 아이템을 추천할 수 있는데, 이용자 프로파일을 구축하기 위해서는 아이템에 대한 이용자의 평점(ratings)을 수집해야 한다. 이를 위해 이용자 참여를 유도하는 것이 쉽지 않다. 이는 비단 콘텐츠기반추천시스템뿐만 아니라 이용자 평점을 필요로 하는 대부분의 추천시스템에서 공통적으로 겪는 문제다.

나. 협업필터링

협업필터링(Collaborative Filtering)은 Collaborative Recommendation, Social Filtering 등 여러 가지 명칭으로 불리며, 한 이용자의 취향과 유사한 다른 이용자의 취향을 취합하여 아이템에 대한 선호도를 예측하는 시스템이다. 협업필터링은 사람들의 취향은 무작위로 분포하는 것이 아니라 일반적인 경향과 패턴이 있다는 가정에서 시작한다. 실제로 사람들은 주위 친구들이나 인터넷을 통해 다른 이용자들과 커뮤니케이션하고 이를 통해 아이템을 추천 받기도 한다. 협업필터링은 이러한 "입소문"(Shardanand 1995)의 개념을 자동화 한 것이라 할 수 있으며 일반적인 절차는 다음과 같다. ① 이용자는 이전에 구매하거나 경험했던 아이템에 대해 평점(rating)을 매기고 이를 통해 이용자 프로파일을 구성한다. ② 같은 아이템에 대해 평점을 내린 이용자들

의 프로파일을 비교하여 유사도에 따라 가중치를 매긴다. 이중 취향이 비슷한 이용자 그룹을 Nearest-Neighborhood라 한다. ③ 마지막으로 어떤 새로운 아이템에 대한 이용자 예상 선호도를 그 아이템을 이미 경험한 Nearest-Neighborhood의 기존 평가를 토대로 예측한다. 협업필터링의 초기 연구 중 하나인 Ringo(Shardanand 1995)는 위의 절차를 통해 아티스트에 대한 이용자 평가를 비교하여 새로운 아티스트들을 추천했다.

협업필터링은 콘텐츠기반추천시스템이 갖고 있는 문제점을 해결해준다. 먼저, 자동으로 분석하기 어려웠던 아이템의 속성들 즉, 영상, 음향, 아이디어, 감정 같은 것을 이용자 평점으로 계산할 수 있기 때문에, 콘텐츠기반추천시스템이 다루지 못했던 아이템 추천이 가능하다. 또한 이용자의 취향이나 아이템의 질에 기반한 추천이 가능하다. 그리고 다른 이용자의 경험을 바탕으로 하기 때문에 평소 선호했던 아이템과는 다르지만, 높이 평가할 수 있는 아이템에 대한 추천도 가능하다. 그러나 협업필터링도 몇 가지 한계점을 가지고 있다. 먼저 새로운 아이템이 추가되었을 때, 이에 대한 이용자 평점이 쌓이기 전에는 이 아이템을 추천할 수 없다. 아울러 아이템 수에 비해 이용자 수가 적을 경우 평점이 존재하지 않는 아이템이 많을 수 있으며, 게다가 이용자가 적으면 공통된 아이템에 대해 평점을 내린 이용자가 작기 때문에 Nearest-Neighborhood를 찾기 어렵다. 특히, 독특한 취향을 가진 이용자의 경우, 유사 취향의 이용자가 드물다면 이 이용자에게 좋은 추천 서비스를 할 수 없다(안신현 2007). 하지만 협업필터링은 추천시스템 분야에서 가장 성공적인 기법으로 전자상거래 기업에서 가장 널리 이용되고 있다(Konstan 1997).

다. Hybrid Recommendation

콘텐츠기반추천시스템과 협업필터링은 서로 상반되는 개념이 아니라 오히려 서로의 단점을 보완함으로써 추천시스템의 성능을 높일 수 있다. Balabanovic(1997)은 웹페이지 추천시스템으로 콘텐츠기반추천시스템과 협업필터링을 접목하여 이용자마다 프로파일을 구축하고 프로파일과 비슷한 아이템을 추천하거나 이웃 이용자가 높게 평가한 아이템을 추천하는 모델을 제시하였다. 콘텐츠기반추천시스템의 특성 때문에 아직 어떤 이용자도 평가하지 않은 아이템을 추천하는 것이 가능하고, 협업필터링의 특성으로 새로운 아이템을 접할 수 있는 기회를 제공할 수 있다.

3.2 나이브베이즈모델

문서분류란 많은 양의 서로 다른 문서들을 미리 정의된 여러 가지 카테고리 중 하나에 속하도록 분류하는 것을 말한다. 문서분류 과정은 기존에 분류된 다양한 카테고리에 문서들을 학습시킨 다음 그 학습정보를 기반으로 아직 분류되지 않은 입력문서를 분류한다. 이러

한 문서분류의 과정에서 필수적인 과정이 문서를 수치로 표현하는 것이다. 일반적으로 문서 수치화는 문서에 포함된 단어를 기준으로 하며, 단어 출현빈도를 기반으로 벡터를 만들어 문서를 분류한다. 학습기반 문서분류 방법에는 Naive Bayes Classifier, Support Vector Machine(SVM), K-Nearest Neighbor (K-NN), Decision Tree, 신경망 등 여러 가지가 있다.

가. 나이브베이즈모델 개관

나이브베이즈모델(Naive Bayesian Classifier)은 문서분류에서 가장 보편적으로 사용하는 방식이다. 베이즈 정리를 이용하여 개발된 이 알고리즘은 텍스트 분류에서 신경망이나 결정트리 학습과 비슷한 성능을 보여주며 자료량이 많아질수록 정확도가 높다. 기본적인 아이디어는 주어진 문서를 입력 받은 뒤 그것이 각 카테고리에 할당될 확률을 계산하는 방법이다. 문서가 특정 카테고리에 속하는 확률을 계산하기 위하여 식 (1)과 (2)와 같이 베이즈 정리를 이용한다.

$$P(c|x) = \frac{P(c)P(x|c)}{P(x)} \quad (1)$$

$$P(x) = \sum_{c \in C} P(c)P(x|c) \quad (2)$$

여기서 x는 임의의 문서이며 c는 임의의 카테고리를 의미한다. 식(1)의 P(x)는 전확률공식(total probability formula)에 의해 식 (2)와 같이 재정의된다. 그런데 P(x)는 모든 카테고리에 대하여 같은 값을 가지므로 확률을 계산하는데 고려하지 않아도 된다. 따라서 식 (1)의 분자에 위치한 P(c)와 P(x|c)만 추정하면 문서 x가 카테고리 c에 할당될 확률을 계산할 수 있다. P(c)는 모든 카테고리 중 카테고리 c가 뽑힐 확률이다. 나이브베이즈모델은 스팸메일처리에서 많이 활용하고 있다. 스팸으로 처리한 메일을 기준으로 앞으로 들어오는 문서가 스팸인지 아닌지를 판단하는 방식이다. 알고리즘 앞에 naive라는 이름이 붙은 이유는 문서 A에 a, b, c라는 단어가 있다고 할 때 b가 나타날 확률은 a나 c와 무관한 즉 독립적이라 가정하여 베이즈 정리를 대폭 단순화한 것이다. 나이브베이즈모델은 일반적으로 문서분류 도메인에서 다른 알고리즘(의사결정 트리, k-nearest neighbor, 신경망 등)에 비해 성능이 우수하다고 알려져 있다(Pedro Domingos 1996). 또한 문서 분류함수를 만드는 것이 간단하고 문서분류 속도도 상대적으로 빨라 문서분류시스템에 많이 이용하고 있다 (김명찬 2003).

나. 나이브베이즈모델 활용

정영미와 이용구(2005)는 문헌 내 단어에서 중의성 문제를 해소하여 검색 성능을 향상시키기 위해 나이브베이즈 분류기를 적용한 실험을 수행하였다. 실험 결과 나이브베이즈 분류기를 적용한 실험에서 92%의 정확률을 나타내며 중의성 문제를 해소하였다. 또한 중

의성 해소를 통한 의미기반 검색 성능은 실험 알고리즘을 적용하지 않은 경우보다 상대적인 정확률에서 7.4% 향상을 보였다.

김판준과 이재윤(2007)은 미분류 문헌에 대한 분류 성능을 향상시키기 위해 문헌 유사도를 자질로 사용하고 나이브베이즈 분류기와 지지벡터기계(Support Vector Machine, SVM)를 이용한 실험을 수행했다. 이 실험은 미분류 문헌을 이용한 자동분류 성능 향상을 평가하는 것이 주된 목적으로 두 가지 분류 모델 모두에서 성능 향상이 나타났다. 이 경우에 SVM이 나이브베이즈 분류기보다 성능 향상 폭이 크게 나타났다.

최근 웹 2.0 환경에서 나이브베이즈 분류 모델을 사용하여 검색 성능 향상과 방법론적으로 상대적인 비교 우위를 검증한 연구들이 있었다. 웹에서 몇 년간 많은 양의 콘텐츠가 축적된 질문-답변 문서에 대한 검색 성능을 향상시키는 연구에서 나이브베이즈 분류 모델이 사용되어 검색 성능 향상을 검증했다(연종흠, 심준호, 이상구 2010). 다음으로 소셜 북마킹 시스템에서 사용자가 입력한 태그에 대한 데이터를 추출하여 사용자가 스패머인지 아닌지를 예측하는 모델에 나이브베이즈 분류기가 사용되어 다른 방법보다 상대적으로 높은 성능을 보였다. 비교를 하기 위해 사용된 방법들은 결정테이블(decision table, DT), 결정트리(decision tree, ID3), TAN(tree-augmented naive Bayes) 분류기, 인공신경망(Artificial Neutral Network) 등이었다. 나이브베이즈 분류기를 사용한 모델이 모델 생성 시간과 AUC(Area Under the ROC Curve)에서 가장 높은 성능을 보였다(김찬주, 황규백 2009).

다. 나이브베이즈모델을 적용한 협업 필터링

협업 필터링에 나이브 베이즈 분류기를 적용한 비교적 초기 연구에서 기존 상관관계 기반 협업 필터링 모델과 비교하여 더 높은 성능을 발휘하는 것을 검증하였다(Koji Miyahara, Michael J. Pazzani 2000). 여기에서는 협업 필터링을 이용자 유사성에 기반한 모델과 아이템 유사성에 기반한 모델 두 가지에 걸쳐 적용하였다. 실험 결과 나이브베이즈 분류기를 적용한 각 협업 필터링 모델이 기존에 상관관계를 이용한 협업 필터링보다 높은 정확률을 나타냈으며 또한 이용자 기반 모델과 아이템 기반 모델을 혼합하였을 경우 더 높은 정확률을 나타냈다.

나이브베이즈 분류기를 적용한 협업 필터링을 향상시키기 위한 연구로는 구간 추정을 이용한 연구가 수행된 바 있다(Robles et al. 2003). 이 연구에서는 UCI 저장소에 있는 마이크로소프트 익명 웹 데이터를 이용하여 기존 모델보다 성능이 향상됨을 증명했다.

4. 제안모델

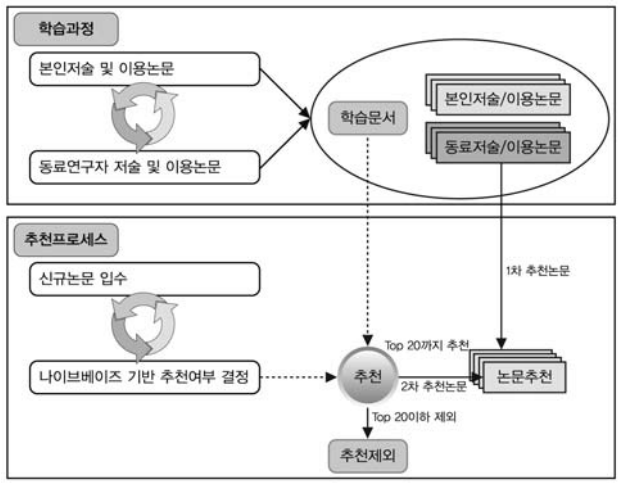
4.1 학습과정 및 추천프로세스

본 연구에서는 대규모 정보센터나 도서관에

서 학술논문을 지능적으로 추천하기 위해 협업필터링 방식에 나이브베이즈모델을 결합한 하이브리드 추천시스템을 제시하였으며, 학습과정과 추천프로세스는 〈그림 3〉과 같다. 먼저 학습과정은 본인이 저술한 논문이나 열람한 논문을 학술논문 DB와 로그파일을 통해 주기적으로 수집하여 초기 학습문서를 구축한다. 그리고 같은 논문을 열람한 관련분야 동료연구자나 공저자 관계에 있는 연구자들이 저술한 논문 또는 열람한 논문을 수집하여 학습문서에 추가한다. 이때 동료 연구자들이 저술한 논문이나 열람한 논문은 협업필터링 방식에 의거 추천대상 논문이 된다. 같은 문서를 열람한 횟수가 많은 동료 연구자나 공저자 빈도에 따라 가중치를 달리 부여하고 이를 토대로 추천문서 우선순위를 산정하였다.

학술논문의 경우 최신성 및 적시성이 매우 중요하고 대규모 정보센터나 도서관의 경우 정보량이 방대하기 때문에 전략적인 접근이 필요하다. 본 연구에서는 효율성을 고려하여 이전에 입수한 학술논문은 문서학습이나 협업필터링 추천에 활용하고 나이브베이즈모델은 신규논문 추천에만 적용하였다. 신규논문이 입수되면 협업필터링에 의해 이미 학습된 연구자 클래스의 자질을 나이브베이즈모델을 이용하여 유사확률을 계산한 후 상위 20편의 논문을 추천하도록 설계하였다. 추천논문은 학습문서에 즉시 추가되는 것은 아니며 추천한 논문을 해당 연구자가 열람할 때 비로소 학습문서에 추가된다. 이용자별 학습문서(클래스)는 본인이 저술하거나 열람한 논문, 공저자 관계에 있거나 같은 논문을 열람한 동료 연구자가 저술 또는 열람한 논문으로 자동구축하고, 자질(feature)은 학습논문의 키워드 항목에서 추출하였다.

본인이 저술하거나 열람한 논문과 동료 연구



〈그림 3〉 학습과정 및 추천프로세스

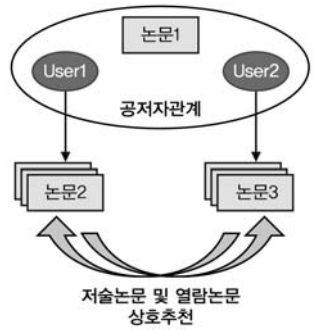
자들이 저술하거나 열람한 학습논문을 같은 비중으로 다루지 않고 중요도에 따라 가중치를 달리 부여하였다. 즉, 본인이 제1저자인 학술논문에 최고점수인 4점, 공저자 또는 동료 연구자들의 저술 논문에 3점, 본인이 열람한 논문에 2점, 동료 연구자가 열람한 논문에 1점을 부여하였다. 아울러 동료 연구자라도 여러 개의 논문을 같이 열람했거나 공저자 빈도가 높을수록 가중치를 달리하였다. 즉, User1과 User2의 공저자 회수가 2회인 경우 x2를 하여 가중치를 배가하였다. 이렇게 부여한 가중치는 협업필터링에 의한 추천논문 순위결정이나 나이브베이즈모델의 자질선정에 큰 영향을 미친다.

4.2 협업필터링 기반 학습과정 및 추천프로세스

가. 공저자에 의한 논문추천 및 학습과정

공저자에 의한 추천과정은 〈그림 4〉와 같이 User1과 User2가 논문1을 공동으로 저술한 경우 User1과 User2는 공저자 관계에 있으며, 이들이 저술한 논문이나 이후 열람한 논문은 상호추천 대상이 된다.

학습과정은 〈그림 5〉와 같이 User1은 공저자 관계에 있는 User2가 저술한 논문3과 열람한 논문5를 학습하고, 반대로 User2는 User1이 저술한 논문2와 열람한 논문4를 학습한다. 이들 논문은 동시에 추천논문이 되며, 추천품질 및 우선순위를 고려하여 중요도에 따라 가중치를 달리하였다. 즉, User1이 제1저자인 논문1에는 가중치 4를 부여했으며, 같은 논문



〈그림 4〉 공저자에 의한 추천과정

class			학습관계	가중치
이용자	학술논문	키워드		
user1	논문1	A, B, C	저작논문(제1저자)	4
	논문2	D, E, F	User1 저작논문(제1저자)	4
	논문4	A, H, I	User1 열람논문	2
	논문5	A, Q, R	User2 열람논문	1
	논문3	P, Q, R	User2 저작논문	3
user2	논문1	A, B, C	저작논문(제2저자 이하)	3
	논문3	P, Q, R	User2 저작논문(제1저자)	4
	논문5	A, Q, R	User2 열람논문	2
	논문4	A, H, I	User1 열람논문	1
	논문2	D, E, F	User1 저작논문	3

상호학습

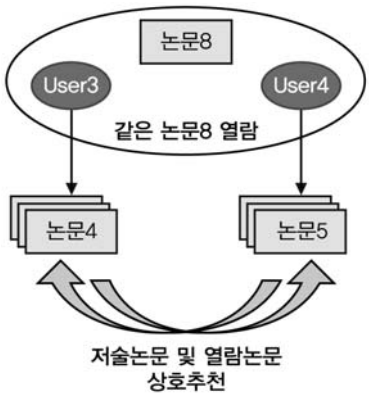
공저자 관계

〈그림 5〉 공저자에 의한 학습과정

이지만 User2의 경우 제2저자인 관계로 가중치 3을 주었다. 또한 본인이 열람한 논문인 User1의 논문4와 User2의 논문5는 가중치가 2인 반면, 상대방이 열람한 논문은 최저 수준인 1을 부여하였다.

나. 동료 연구자에 의한 논문추천 및 학습과정

〈그림 6〉은 동료 연구자에 의한 논문추천 및 학습과정을 도식화한 것이다. User3과 User4가 논문8을 같이 열람한 경우, User3과 User4는 연구 분야가 유사한 동료 연구자로 볼 수 있으므로 이들이 저술하거나 열람한 논문을 상호 추천하였다.



〈그림 6〉 동료 연구자에 의한 추천과정

동료 연구자에 의한 학습과정은 〈그림 7〉과 같이 User3은 자신이 저술하거나 열람한 논문2, 4, 6을 User4에게 학습하고, User4는 자신이 저술하거나 열람한 논문3, 5, 7을 User3에게 학습한다. 이들 논문은 이후 상호 추천논문이 되며 공저자 관계에서와 마찬가지로 중요도에 따라 가중치를 달리하였다.

4.3 나이브베이즈모델 기반 신규논문 추천 프로세스

가. 나이브베이즈모델 유사확률 계산 알고리즘

본 연구에 적용한 나이브베이즈모델의 유사확률 계산 알고리즘은 다음과 같다. 나이브베이즈모델에서 문서 D가 C클래스일 확률은 $P(D) = P(t1, t2, \cdots, tn)$, $P(C|D) = P(C|t1, t2, \cdots, tn) = (P(C)*(t1, t2, \cdots, tn|C)) / P(t1, t2, \cdots, tn)$로 정리할 수 있다. 예를 들

class			학습관계	가중치	
이용자	학술논문	키워드			
User3	논문8	A, B, C	열람논문	2	
	논문2	D, E, F	User3 저작논문(제1저자)	4	
	논문4	A, H, I	User3 저작논문(제2저자 이하)	3	
	논문6	A, H, I	User3 열람논문	2	
User4	논문8	A, B, C	열람논문	2	
	논문3	P, Q, R	User4 저작논문(제1저자...)	4	
	논문5	A, H, I	User4 저작논문(제2저자 이하)	3	
	논문7	A, H, I	User4 열람논문	2	

상호 학습

같은 논문 열람

〈그림 7〉 동료 연구자에 의한 학습과정

면 학습문서 4개가 있고 2개의 클래스로 구성되었다고 가정하면(A, B, …, F는 각 문서의 키워드) 다음과 같이 나이브베이즈 확률을 계산할 수 있다.

(User1) 문서1 : A, B, A

(User1) 문서2 : A, A, C

(User1) 문서3 : A, D

(User2) 문서4 : F, E, A

각 키워드의 확률을 계산하면

P(A | User1) = (User1에서 A의 출현횟수 / User1의 전체 단어수) = 5/8

P(B | User1) = 1/8

P(C | User1) = 1/8

P(D | User1) = 1/8

P(E | User1) = 0/8

P(F | User1) = 0/8

P(A | User2) = (User2에서 A의 출현횟수 / User2의 전체 단어수) = 1/3

P(B | User2) = 0/3

P(C | User2) = 0/3

P(D | User2) = 0/3

P(E | User2) = 1/3

P(F | User2) = 1/3

P(User1) = 3/4

P(User2) = 1/4이다.

만약, 신규문서가 A, A, A, F, E라는 키워드로 구성되었다고 가정하면

P(User1 | 신규문서) = 3/4 * 5/8 * 5/8 * 5/8* 0/8 * 0/8

P(User2 | 신규문서) = 1/4 * 1/3 * 1/3 * 1/3* 1/3 * 1/3이 된다.

그런데 위에서와 같이 빈도수가 0이면 이 값으로 인해 모든 확률값이 0이 되므로 본 논문에서는 이를 방지하기 위해 분자에 +1을 하고 분모에 전체 학습문서 자질수를 +하였다. 예를 들면 P(E | 클래스1) = {(클래스1에서 E의 출현횟수 + 1) / (클래스1의 전체 키워드 + 학습문서 자질 수)} = (0+1)/(8+6)로 보정하였다. 이를 적용하면 다음과 같다.

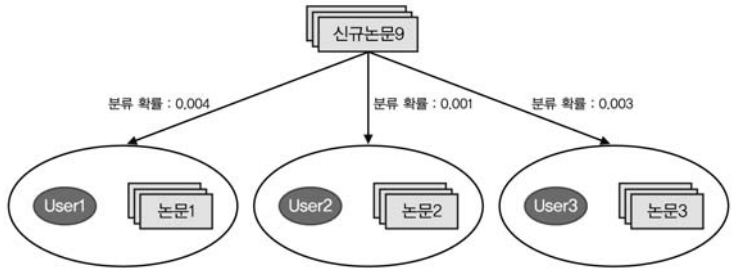
P(User1 | 신규문서) = 3/4 * 5/14 * 5/14 * 5/14 * 1/14 * 1/14 = 0.0001743135

P(User2 | 신규문서) = 1/4 * 1/9 * 1/9 * 1/9 * 1/9 * 1/9 = 0.0000042338로 신규문서는 User2보다 User1에 훨씬 더 유사하다는 것을 알 수 있다.

나. 신규논문 추천 프로세스

대규모 정보센터나 도서관의 경우 정보량이 방대하기 때문에 과거문서까지 소급하여 추천하는 것은 매우 비효율적이다. 본 논문에서는 최신성과 적시성이 중요한 학술논문의 특성을 고려하여 신규로 입수하는 학술논문에만 나이브베이즈모델을 적용하였으며, 학습문서 및 자질은 협업필터링 방식을 통해 지속적으로 자동 구축되도록 설계하였다.

신규논문이 입수되면 나이브베이즈모델로 확률값을 계산하고 각 이용자별 상위 20개 논문을 추천한다. 즉 〈그림 8〉에서와 같이 신규논문9가 입수되면 나이브베이즈모델로 확률값



〈그림 8〉 나이브베이즈모델 기반 신규논문 추천과정

을 계산하여 어떤 이용자에게 추천하는 것이 적정한지를 판단하고, 추천한 논문을 이용자가 열람하면 그 시점에서 자동적으로 학습문서에 포함되도록 모델링하였다.

5. 실험결과

본 모델을 검증하기위해 한국과학기술정보연구원의 NDSL 학술논문과 이용로그를 활용하여 실험하였다. 한국과학기술정보연구원에는 현재(2010.8) 5천만여 건의 해외학술논문과 1백만여 건의 국내학술논문이 구축되어 있으며, 60만여 명의 회원이 이를 이용하고 있다. 실험에 사용된 데이터는 최근 3년간(2007~2009)의 NDSL 이용로그를 분석하여 이용 빈도가 높은 학술논문 약 3천 5백 건을 먼저 추출하고, 이를 많이 이용한 식품과 전기 분야 연구자 상위 5명을 선정하였다.

〈표 1〉은 이용 빈도가 높은 상위 3천 5백

〈표 1〉 이용 빈도가 높은 학술논문 순위

A	B	C	D	E	F	G	I	J
ARTCTRLNO	TITLE	JOURNALNAME	PUBLISHER	VOLUM/ISSUE	AUTHOR	KEYWORD	낳은논문열람자명	EMAIL
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	최철웅	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	최봉규	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	전봉주	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	장정화	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	김은영	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	황용철	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	문지영	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	최유석	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	정준호	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	박가현	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	허병석	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	조영낭	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	김선오	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	박선영	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	서진숙	
34511584	홍화씨추출물의	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	김미진 ; 김자영	Phytoestrogen	박성욱	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	방미애	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	박찬하	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	문영건	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	장미현	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	이재연	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	강경희	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	김기영	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	김민지	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	배준태	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	박상재	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	박현서	
34511597	아토피 피부염	대한화장품학회	대한화장품학회	v.30 no.4 = no.4	유창선 ; 김선희	atopic dermatitis	황웅주	

편의 학술논문 목록이며, 본 논문의 연구자 식별에 이용하는 e-mail 등은 보안상 비공개 처리하였다.

5.1 협업필터링에 의한 실험결과

〈표 2〉와 같이 실험 대상자 5명 중 한명인 강○○ 연구자는 식품분야 국내논문 2편, “향토주인 산성막걸리의 미생물학적 고찰과 저장성에 관한 연구”와 “각종 누룩 균을 이용한 실험실 조건에서의 막걸리 제조”를 열람하였으며, 같은 논문을 열람한 동료 연구자는 각각 15명과 16명으로 나타났다.

동일한 논문을 열람한 연구자들은 연구 분야가 유사한 동료 연구자일 확률이 높으므로 이들이 저술하거나 열람한 논문을 상호 추천하였다. 즉 박○○ 연구자의 경우 총3편의 논문을 열람하였으며 강○ 연구자와 같이 열람한 “향토주인 산성막걸리의 미생물학적 고찰과 저장성에 관한 연구”를 제외하고 박○○ 연구자만 열람한 2편의 논문, “2단 담금에서 무증자 발아현미를 이용한 막걸리 제조”와 “누룩에서 분리한 효모를 이용한 찹쌀발효주의 이화학적 특성 및 휘발성 향기성분”을 강○○ 연구자에게 추천하였다.

이에 비해 송○○ 연구자의 경우 강○○ 연구자와 2편의 같은 논문을 열람하여, 박○○에 비해 훨씬 연구 분야가 밀접하므로 〈표 4〉에서와 같이 가중치를 달리 부여하여 송○○ 연구자가 저술하거나 열람한 논문의 추천 우

〈표 2〉 같은 논문 열람자

TITLE	JOURNALNAME	VOLUMEN	AUTHOR	KEYWORD	같은논문열람자명
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	강ㅇㅇ
각종 누룩 균을 이용	한국균학회지	v.30 no.1	이상선 ; 김교숙	Absidia . Aspergil	강ㅇㅇ
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	박
2단담금에서 무증자	한국식품영양과학	v.32 no.6	송재철 ; 박현정	uncooked germed	박
누룩에서 분리한 효모	한국식품과학회지	v.41 no.3	김혜련 ; 권영희	volatile flavor com	박
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	최
2단담금에서 무증자	한국식품영양과학	v.32 no.6	송재철 ; 박현정	uncooked germed	최
막걸리 분획물에 의	한국식품영양과학	v.37 no.3	신미옥 ; 강대연	cytotoxicity . quin	최
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	him
각종 누룩 균을 이용	한국균학회지	v.30 no.1	이상선 ; 김교숙	Absidia . Aspergil	him
오디(Morus alba) 와	한국식품과학회지	v.40 no.1	김용석 ; 정도영	mulberry . mulber	him
고효율 바이오 에탄	한국생물공학회지	v.23 no.6	최기욱 ; 한민희	bioethanol . simul	him
어성초 추출물의 발	한국식품저장유통	v.16 no.1	김미림 ; Kim, M	Houttuynia cordat	him
누룩에서 분리한 효모	한국식품과학회지	v.41 no.3	김혜련 ; 권영희	volatile flavor com	him
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	송
각종 누룩 균을 이용	한국균학회지	v.30 no.1	이상선 ; 김교숙	Absidia . Aspergil	송
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	정
누룩에서 분리한 효모	한국식품과학회지	v.41 no.3	김혜련 ; 권영희	volatile flavor com	정
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	대
누룩에서 분리한 효모	한국식품과학회지	v.41 no.3	김혜련 ; 권영희	volatile flavor com	대
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	이
각종 누룩 균을 이용	한국균학회지	v.30 no.1	이상선 ; 김교숙	Absidia . Aspergil	이
막걸리 분획물에 의	한국식품영양과학	v.37 no.3	신미옥 ; 강대연	cytotoxicity . quin	이
와송 부위별 추출물	한국식품영양과학	v.38 no.1	윤소영 ; 이소영	Orostachys japon	이
향토주인 산성막걸리	한국식품과학회지	v.28 no.4	양지영 ; 이계호	Sansung takju . k	서
어성초(魚腥草) 추출	대한약침학회지	v.11 no.1	전오도 ; 서형식	Houttuyniae Herb	서
어성초 ethanol 추출	한국환경농학회지	v.27 no.2	김동혁 ; 임정주	Salmonellosis . H	서
어성초 추출물의 발	한국식품저장유통	v.16 no.1	김미림 ; Kim, M	Houttuynia cordat	서

선순위를 산정하였다.

〈표 3〉은 가중치를 전혀 고려하지 않고 강○○에게 추천한 논문 우선순위로, 강○○가 열람하거나 저술한 논문을 동료 연구자들이 얼마나 많이 열람했는지? 열람 횟수를 토대로 추천한 논문 우선순위이며, 〈표 4〉는 가중치를 적용한 추천논문 우선순위이다. 〈표 3〉과 〈표 4〉의 추천논문 우선순위를 비교해 보면, 가중치를 적용한 경우와 적용하지 않은 경우 추천논문 우선순위에 많은 변화가 있는 것을 알 수 있다. 추천시스템 목적이 대량의 정보자원 중에서 연구자에게 적합한 논문을 예측하여 제공하는 것이므로 우선순위는 매우 중요한 요소이다.

가중치를 적용한 추천논문과 그렇지 않은 추천논문을 실험대상자 5명에게 이메일로 발

〈표 3〉 가중치 미적용 추천논문 순위

TITLE	JOURNALNAME	VOLUMENA	AUTHOR	KEYWORD	이용자 수
2단담금에서 무증자 발아현미를 이용한 막걸리 제	한국식품영양과학회	v.32 no.6	송재철 ; 박현정	uncooked germed br	7
누룩에서 분리한 효모를 이용한 찹쌀발효주의 이호	한국식품과학회지	v.41 no.3	김혜련 ; 권영희 ; 조	volatile flavor compor	7
막걸리 분획물에 의한 암세포 성장 억제 및 Quinor	한국식품영양과학회	v.37 no.3	신미옥 ; 강대연 ; 김	cytotoxicity . quinon	6
항염증 활성이 증진된 민들레 잎의 열수 추출물 제	한국식품영양과학회	v.38 no.3	고윤정 ; 박용곤 ; 김	dandelion leaves . ar	2
한방원료의 초임계 추출을 이용한 항노화 및 주름개	한국생물공학회지	v.23 no.6	김인덕 ; 권륜희 ; 허	Oriental Herb . super	2
어성초 추출물의 발효적성 및 발효액의 특성 비교	한국식품저장유통학	v.16 no.1	김미림 ; Kim, Mi-Lim	Houttuynia cordata T	2
피부 미용을 위한 기능성 식이 소재의 이해	대한화장품학회지	v.30 no.3 = n	김주영 ; 김현애 ; 박	skin . functional food	2
불가사리 유래 콜라겐 펩타이드의 피부 면역 증진	한국식품과학회지	v.40 no.5	정향숙 ; 권민철 ; 한	Asterias amurensis .	2
발효 현미식초의 발효방법 및 원료함량에 따른 품	한국식품저장유통학	v.16 no.1	주경호 ; 조명휘 ; 박	brown rice . Nuruk . v	2
용매 종류와 효소 처리에 따른 쓴 메밀 추출물의 항	한국식품영양과학회	v.38 no.8	김지은 ; 주성일 ; 서	Fagopyrum tataricum	1
약용식물의 열수추출물과 적정 조성추출물 및 그	한국식품저장유통학	v.16 no.2	이가순 ; 김관후 ; 성	medicinal herbs . hot	1
쌀 품종별 백미와 현미의 영양성분 조성 비교	한국식품영양과학회	v.31 no.5	최정숙 ; 안훈희 ; 남	rice . brown rice . nu	1
추출조건에 따른 홍삼추출액의 사포닌 조성 및 이	한국식품영양과학회	v.37 no.2	이승호 ; 강정일 ; 이	Korean red ginseng .	1
HPLC를 이용한 인삼, 홍삼, 산양산삼 및 홍산삼의	대한약침학회지	v.11 no.2	이장호 ; 권기록 ; 차	cultivated ginseng . r	1
흑마늘 가공 중 이화학적 성분 및 항산화 활성의 변	생명과학회지	v.18 no.8	신정혜 ; 최덕주 ; 이	Black garlic . free su	1
인삼 및 산양삼의 항산화 효능 비교 -Superoxide ra	대한약침학회지	v.12 no.2	임태진 ; 정희선 ; 김	antioxidant activity .	1
어성초(魚腥草) 추출물의 항여드름 효과에 관한 연	대한약침학회지	v.11 no.1	전오도 ; 서형식 ; Jec	Houttuyniae Herba .	1
어성초 ethanol 추출물의 마우스 살모넬라 감염증에	한국환경농학회지	v.27 no.2	김동혁 ; 임정주 ; 이	Salmonellosis . Hout	1
노니 추출물의 주름개선 효과연구	대한화장품학회지	v.32 no.4 = n	이정노 ; 김상우 ; 유	anti-wrinkle . noni ex	1
와송 부위별 추출물의 항균활성	한국식품영양과학회	v.38 no.1	윤소영 ; 이소영 ; 김	Orostachys japonicu	1

〈표 4〉 가중치 적용 추천논문 순위

TITLE	JOURNALNAME	OLUMENAM	AUTHOR	KEYWORD	가중치	가중치 합
막걸리 분획물에 의한 암세포 성장 억제 및 Qu	한국식품영양과학회	v.37 no.3	신미옥 ; 강대연	cytotoxicity . quin	1	10
2단담금에서 무증자 발아현미를 이용한 막걸리	한국식품영양과학회	v.32 no.6	송재철 ; 박현정	uncooked germed	2	8
누룩에서 분리한 효모를 이용한 찹쌀발효주의	한국식품과학회지	v.41 no.3	김혜련 ; 권영희	volatile flavor comp	1	8
항염증 활성이 증진된 민들레 잎의 열수 추출물	한국식품영양과학회	v.38 no.3	고윤정 ; 박용곤	dandelion leaves .	1	3
어성초 추출물의 발효적성 및 발효액의 특성 비	한국식품저장유통학	v.16 no.1	김미림 ; Kim, M	Houttuynia cordata	1	3
피부 미용을 위한 기능성 식이 소재의 이해	대한화장품학회지	v.30 no.3 =	김주영 ; 김현애	skin . functional fo	1	3
발효 현미식초의 발효방법 및 원료함량에 따른	한국식품저장유통학	v.16 no.1	주경호 ; 조명휘	brown rice . Nuruk	1	3
한방원료의 초임계 추출을 이용한 항노화 및 주	한국생물공학회지	v.23 no.6	김인덕 ; 권륜희	Oriental Herb . sup	1	2
불가사리 유래 콜라겐 펩타이드의 피부 면역 증	한국식품과학회지	v.40 no.5	정향숙 ; 권민철	Asterias amurensi	1	2
와송 부위별 추출물의 항균활성	한국식품영양과학회	v.38 no.1	윤소영 ; 이소영	Orostachys japoni	2	2
홍화씨추출물의 피부 주름개선 효과(II)	대한화장품학회지	v.30 no.4 =	김미진 ; 김자영	Phytoestrogen . s	2	2
종류별 커피의 볶음 및 추출조건에 따른 품질 특	한국식품영양과학회	v.20 no.1	김하경 ; 황성연	coffee . green bea	2	2
가열처리 및 허브첨가에 의한 무취 마늘 소재 개	한국식품영양과학회	v.38 no.1	전미라 ; 김민희	garlic . green tea .	2	2
반응표면분석을 이용한 트레할로스와 변성전분	한국식품영양과학회	v.38 no.3	김상숙 ; 정혜영	trehalose . Sun-Te	2	2
학령 전 아동의 식습관, 신체 발달 및 영양 섭취	한국영양학회지	v.42 no.1	유경희 ; Yu, Ky	preschool children	2	2
현미 종류별 발아현미의 발아 전 후 생리활성물	한국작물학회지	v.53 no.sp	정혜영 ; 이동현	germinated brown	2	2
콜라겐 펩타이드의 피부 광노화 예방 효과	한국식품과학회지	v.41 no.4	김정기 ; 이지해	AP-CPM01 . colla	2	2
오디(Morus alba) 와인의 최적 발효조건 및 발효	한국식품과학회지	v.40 no.1 =	김용석 ; 정도영	mulberry . mulberr	2	2
고효율 바이오 에탄올 생산을 위한 당화효소 가	한국생물공학회지	v.23 no.6	최기욱 ; 한민희	bioethanol . simult	2	2
용매 종류와 효소 처리에 따른 쓴 메밀 추출물의	한국식품영양과학회	v.38 no.8	김지은 ; 주성일	Fagopyrum tataric	1	1

송하고 조사한 결과, 참여자 모두 가중치를 적용하지 않은 것보다 가중치를 적용한 추천논문을 더 선호하는 것으로 나타났다.

5.2 나이브베이즈모델 기반 신규논문 추천결과

나이브베이즈모델을 이용한 신규논문 추천품질도 협업필터링과 동일하게 식품과 전기분야 연구자 5명을 대상으로 실시하였다. 식품과 전기 분야 연구자로 나누어 선정한 이유는 연구영역이 확연히 구분되는 두 개 도메인을 비교하기 위해서다. 실험에 이용한 논문은 금년도(2010) 6월까지 NDSL에서 신규 구축한 국내학술논문 중 5명의 실험 대상자와 관련성이 높은 11개 학술지(한국식품영양과학회지, 한국식품과학회지, 한국식품저장유통학회지, 대한화장품학회지, 한국조리과학회지, 한국전기화학회, 대한전자공학회, 한국자동차공학회, 대한기계학회, 한국전기화학회, 한국전기전자재료학회)에 수록된 980건의 논문을 대상으로 실시하였다.

〈표 5〉는 연구 분야가 식품인 강○○의 학습 클래스로 총71편의 논문을 학습하였다.

〈표 5〉 강OO 학습논문

ARTCTRLNO	TITLE	JOURNALNAME	PUBLISHER	VOLUMENAME	KEYWORD	가중치	가중치 구분
JAKO200825457	한국 전통포도주의	한국식품영양과학	The Korean Socie	v.37 no.8	Ko . rean traditional grape wine . gluti	4	제1저자
JAKO199703042	액체홍국코지를 이	한국식품과학회지	한국식품과학회	v.29 no.1	fermented hot pepper soybean paste	4	제1저자
JAKO199503042	무화과잎 페놀성	한국농화학회지 =	Tong Hakhoe	v.38 no.4	무화과잎 항미생물 활성물질 페놀	4	제1저자
JAKO200111920	무화과를 이용한	한국식품영양과학	The Korean Socie	v.30 no.6	anchovy sauce . fig . fermentation	4	제1저자
JAKO199511920	액체 홍국의 배양	산업미생물학회지	한국산업미생물학	v.23 no.4	Monascus anka . production of red pi	4	제1저자
JAKO199903042	김분말의 제조와	한국식품과학회지	한국식품과학회	v.31 no.5	laver . powder . laver pigment	3	제2저자
JAKO200003042	누룩과 효모의 혼	한국식품과학회지	한국식품과학회	v.32 no.5	nuruk . saccharomyces bayanus . ho	3	제2저자
JAKO200003042	소금의 종류별 무	한국식품과학회지	한국식품과학회	v.32 no.6	sea salt . mineral content . heavy me	3	제2저자
JAKO200009409	1개 초등학교 3학	대한물리치료학회	대한물리치료학회	v.12 no.1		3	제2저자
JAKO200411922	약용식물 추출물에	한국식품저장유통	한국식품저장유통	v.11 no.2	medicinal herbs . antimicrobial activit	3	제2저자
JAKO200211920	황소개구리에서 추	한국식품영양과학	The Korean Socie	v.31 no.2	bullfrog . docosahexaenoic acid . eic	3	제2저자
JAKO200003042	김분말을 이용한	한국식품과학회지	한국식품과학회	v.32 no.2	laver noodle . composite flour . senso	3	제2저자
JAKO200304637	변색김으로부터 p	한국식품과학회지	한국식품과학회	v.35 no.6	Porphyra yezoensis . dried laver decc	3	제2저자
JAKO200918839	고추장 숙성 전복	한국식품영양과학	The Korean Socie	v.38 no.6	Kochujang . cooked abalone . aged a	3	제2저자
JAKO200504637	페놀성 화합물의	韓國食生活文化學	Korean Society of	v.20 no.5	phenolic compounds . anticariogenic	3	제2저자
JAKO199811920	황색고구마 퓨레의	한국식품영양과학	한국식품영양과학	v.27 no.4	yellow sweet potato puree . Benihaya	3	제2저자
JAKO200411922	미역에 함유된 Fu	한국식품영양과학	The Korean Socie	v.33 no.5	Undaria pinnatifida . fucoxanthin . aut	3	제2저자
JAKO200411922	양파껍질 추출물이	한국식품영양과학	The Korean Socie	v.33 no.8	Gulbi . Yellow croaker . brine salting	3	제2저자
JAKO198811922	콩과 옥수수 가루	산업미생물학회지	한국산업미생물학	v.16 no.6	Phytase-producing Microorganism . F	3	제2저자
JAKO199903042	건조김 제조시 키	한국식품과학회지	한국식품과학회	v.31 no.4	chitosan . laver . rice rolled with laver	3	제2저자
JAKO200511722	알긴산을 이용한	한국식품저장유통	한국식품저장유통	v.12 no.6	alginic acid . coated rice . retrogradat	3	제2저자
JAKO199503534	카라기난 생고분자	한국포장학회지 =	한국포장학회	v.1 no.1		3	제2저자
JAKO199803042	카라기난 필름 및	한국식품과학회지	한국식품과학회	v.30 no.2	lipid penetration . carrageenan-based	3	제2저자
JAKO199915875	양파를 이용한 식	산업미생물학회지	한국산업미생물학	v.27 no.1	onion . onion juice . alcohol fermentat	3	제저2자
JAKO199011920	청국장 제조과정에	산업미생물학회지	한국산업미생물학	v.18 no.4	phytase-producing bacteria . Phytase	3	제2저자
JAKO200908856	건조방법에 따른	한국식품저장유통	한국식품저장유통	v.16 no.5	abalone . shady sun drying . cold air	3	제2저자
JAKO199711923	Jar Fermenter에	한국식품영양과학	한국식품영양과학	v.26 no.1	Monascus anka pigments . natural fo	3	제2저자
JAKO200111920	고추장굴비의 저	한국식품영양과학	The Korean Socie	v.30 no.6	traditional food . TBARS valuers . koc	3	제2저자
JAKO200411922	약용식물 추출물에	한국식품저장유통	한국식품저장유통	v.11 no.2	medicinal herbs . DPPH . nitrite-scav	3	제2저자
JAKO200511722	주정을 첨가한 양	한국식품저장유통	한국식품저장유통	v.12 no.6	onion . vinegar . flavonoid . quercetin	3	제2저자
JAKO200520828	연근 분말을 첨가	한국식품영양과학	The Korean Socie	v.34 no.4	soybean paste . lotus root powder . p	3	제2저자
JAKO200411922	다시마, 미역 및 멸	한국식품영양과학	The Korean Socie	v.33 no.5	soybean paste . sweet tangle . sea m	3	제2저자
JAKO200911764	전복(Haliotis disc	한국식품저장유통	한국식품저장유통	v.16 no.6	abalone . amino acid . fatty acid . ch	3	제2저자
JAKO200603042	전복(Haliotis disc	한국식품영양과학	The Korean Socie	v.35 no.7	Haliotis discus hannai . in vitro . angi	3	제2저자
JAKO200411922	미역귀를 이용한	한국식품영양과학	The Korean Socie	v.33 no.3	Undaria pinnatifida suorophyll . jam	3	제2저자
JAKO199903042	고구마의 품종별	한국식품과학회지	한국식품과학회	v.31 no.4	sweet potato . antioxidative activity .	3	제2저자
JAKO200003042	천일염으로 제조한	한국식품과학회지	한국식품과학회	v.32 no.6	sea salts . doenjang . fermentation	3	제2저자
34731623	각종 누룩 균을 이	한국균학회지	한국균학회	v.30 no.1	Absidia . Aspergillus . Korean rice wi	2	본인열람
34519502	향토주인 산성막걸	한국식품과학회지	한국식품과학회	v.28 no.4	Sansung takju . kokja . shelf-life	2	본인열람
36437132	2단담금에서 무증	한국식품영양과학	한국식품영양과학	v.32 no.6	uncooked germed brown rice . Takju	2	중복이용
48872631	와송 부위별 추출	한국식품영양과학	한국식품영양과학	v.38 no.1	Orostachys japonicus . food borne pa	2	중복이용
34511584	홍화씨추출물의 표	대한화장품학회지	대한화장품학회	v.30 no.4 = no.48	Phytoestrogen . safflower (Carthamus	2	중복이용
36970810	종류별 커피의 볶	한국식품영양학회	한국식품영양학회	v.20 no.1	coffee . green bean . roasted bean . e	2	중복이용
48872644	가열처리 및 허브	한국식품영양과학	한국식품영양과학	v.38 no.1	garlic . green tea . heat processing .	2	중복이용
49195237	반응표면분석을 이	한국식품영양과학	한국식품영양과학	v.38 no.3	trehalose . Sun-Tender . texture prop	2	중복이용
49211290	학령 전 아동의 식	한국영양학회지	한국영양학회	v.42 no.1	preschool children . nutrients intakes	2	중복이용
49215671	현미 종류별 발아	한국작물학회지	한국작물학회	v.53 no.spc	germinated brown rice . phytosterol	2	중복이용
50292563	콜라겐 펩타이드의	한국식품과학회지	한국식품과학회	v.41 no.4	AP-CPM01 . collagen peptide . elasti	2	중복이용
47250361	오디(Morus alba)	한국식품과학회지	한국식품과학회	v.40 no.1 = no.191	mulberry . mulberry wine . optimum fe	2	중복이용
49209772	고효율 바이오 에	한국생물공학회지	한국생물공학회	v.23 no.6	bioethanol . simultaneous saccharific	2	중복이용
46984101	막걸리 분획물에	한국식품영양과학	한국식품영양과학	v.37 no.3	cytotoxicity . quinone reductase . Ma	1	추천논문
49970074	누룩에서 분리한	한국식품과학회지	한국식품과학회	v.41 no.3	volatile flavor component . glutinous ri	1	추천논문

즉, 본인이 저술한 논문 4편, 공저자 논문 32편, 본인열람 2편, 동료 추천논문 33편이다.

강○○ 학습논문 71편의 키워드를 추출하여 자질을 구성한 다음 나이브베이즈모델로 신규논문 980편과 유사확률을 계산한 결과, 〈표 6〉과 같이 한국식품영양과학회지 논문이 대부분 상위에 랭크되었으며, 〈표 7〉에서와 같이 같은 저널의 논문이라도 연구 분야와 동떨어지거나 전혀 관련이 없는 한국자동차공학회 논문은 하위에 랭크된 것을 볼 수 있다.

〈표 8〉은 식품분야의 또 다른 실험 대상자인 김○○ 연구자에 대한 나이브베이즈모델 기반 신규 논문의 유사확률로 강○○ 연구자와는 확연히 다른 결과를 보여주고 있다. 이는

〈표 6〉 강OO 상위 유사확률(추천대상)

1	TITLE	JOURNALNAME	VOLUMENAME	AUTHOR	KEYWORDS	강OO 유사확률
2	연잎 에탄올 추출물의 항산화 효과	한국식품영양과학회지 = J	v.35,0 no.2,0	이경석 ; 김민규 ; 0	antioxidant activity , lotus (Nelur	0,0000059259261908354500000000000
3	유자(Citrus junos SIEB ex TANAKA)종실 에탄올추	한국식품영양과학회지 = J	v.35,0 no.3,0	권오천 ; 신정혜 ; 김	citron seed , antioxidant activity	0,0000059259261908354500000000000
4	식중독유발 세균의 증식에 미치는 천초근 추출물	한국식품영양과학회지 = J	v.34,0 no.3,0	배지현 ; 장혜정 ; 정	Rubia akane Nakai , antimicrobia	0,0000000237037058892073000000000
5	연잎 추출물의 항균 효과	한국식품영양과학회지 = J	v.35,0 no.2,0	이경석 ; 오창석 ; 0	lotus (Nelumbo nucifera) leaf , a	0,0000000237037058892073000000000
6	식중독유발 세균의 증식에 미치는 백화사설초 추	한국식품영양과학회지 = J	v.34,0 no.1,0	배지현	Hedyotis diffusa , antimicrobial a	0,0000000237037058892073000000000
7	밀 단백 효소 가수분해물의 항균활성	한국식품영양과학회지 = J	v.32,0 no.5,0	이상덕 ; 주정현 ; 0	wheat protein , protease , antimi	0,0000000237037058892073000000000
8	황금탕 및 황금추출물의 식중독 세균들에 대한	한국식품영양과학회지 = J	v.34,0 no.10,0	서진종 ; 이종기	Whangkumtang extract , Scutella	0,0000000237037058892073000000000
9	채소류 메탄올 추출물의 In Vitro와 Cell System에	한국식품영양과학회지 = J	v.34,0 no.8,0	이영아 ; 김현영 ; 조	antioxidative activity , vegetable	0,0000000164609064842447000000000
10	갈참나무 부위별 에탄올 추출물 및 분획물의 항	한국식품영양과학회지 = J	v.34,0 no.6,0	윤재원 ; 유미영 ; 초	antimicrobial activity , Quercus a	0,0000000219478743925015000000000
11	채취시기에 따른 미더덕의 부위별 항산화 활성	한국식품영양과학회지 = J	v.39,0 no.3,0	이동원 ; 유동현 ; 임	Styela clava , antioxidant activit	0,0000000237037058892073000000000
12	와송(Orostachys japonicus) 잎, 줄기 및 뿌리 추	한국식품영양과학회지 = J	v.38,0 no.11,0	이소정 ; 송유진 ; 0	Orostachys japonicus , antioxida	0,0000000219478743925015000000000
13	칼슘 분말제재 첨가가 김치의 숙성 중 품질에 미	한국식품영양과학회지 = J	v.31,0 no.3,0	박우포 ; 박규동 ; 정	kimchi , fermentation , calcium p	0,0000000533333419760073000000000
14	비트 첨가가 미국인 선호 김치의 숙성 중 품질에	한국식품영양과학회지 = J	v.34,0 no.4,0	양유진 ; 한지숙	American preferred kimchi , beet	0,0000000533333419760073000000000
15	나박김치의 제조 표준화 및 발효온도별 젖산균의	한국식품영양과학회지 = J	v.34,0 no.5,0	공창숙 ; 서정옥 ; 브	nabak kimchi , standardization ,	0,0000000533333419760073000000000
16	Cook/Chill System에서의 고기완자 제조에 관한	한국식품영양과학회지 = J	v.32,0 no.5,0	강은주 ; 김선영 ; 유	Wanjajun , cook/chill system , a	0,0000000118518529446036000000000
17	김 분획물의 in vitro에서의 항발암효과	한국식품영양과학회지 = J	v.34,0 no.10,0	신미옥 ; 배송자	cytotoxicity , quinone reductase	0,0000000118518529446036000000000
18	추출 조건에 따른 어수리의 항산화 활성 및 성분	한국식품저장유통학회지 =	v.16,0 no.5,0	방지은 ; 최해연 ; 김	anti-oxidative activity , Heracleu	0,0000000000000000000000000000000
19	대두박으로부터 분리한 식용 조사포닌의 항암 및	한국식품영양과학회지 = J	v.34,0 no.10,0	박경욱 ; 위재준 ; 김	saponin , anticancer and immune	0,0000000118518529446036000000000
20	불등가사리 분획물의 암세포 성장 억제 효과	한국식품영양과학회지 = J	v.34,0 no.6,0	박성영 ; 정복미 ; 초	cytotoxicity , quinone reductase	0,0000000109739371962507000000000
21	참가사리 분획물의 암 예방효과	한국식품영양과학회지 = J	v.35,0 no.4,0	정영화 ; 정복미 ; 신	cytotoxicity , quinone reductase	0,0000000109739371962507000000000

〈표 7〉 강OO 하위 유사확률(추천제외)

1	TITLE	JOURNALNAME	VOLUMENAME	AUTHOR	KEYWORDS	강OO 유사확률
280	부산지역 고등학교 학생과 급식종사자의 급식위	한국식품영양과학회지 = J	v.34,0 no.9,0	김소희	foodservice , importance , perfo	0,0000000000000000000000000000000
281	부산지역 학교급식 영양사의 이벤트 마케팅에 대	한국식품영양과학회지 = J	v.38,0 no.12,0	이경아	foodservice event , school food	0,0000000000000000000000000000000
282	사염화탄소를 투여한 흰쥐의 간세포에 대한 식물	한국식품영양과학회지 = J	v.33,0 no.8,0	함영국 ; 김성완	plant extracts , <TEX>CCl_4<	0,0000000000000000000000000000000
283	충북지역 유치원 급식품질에 대한 학부모 만족도	한국식품영양과학회지 = J	v.39,0 no.4,0	이주연 ; 이연은	foodservice quality , kindergarte	0,0000000000000000000000000000000
284	부산지역 제과제빵종사자의 위생요인에 대한 중	한국식품영양과학회지 = J	v.39,0 no.4,0	김소희	hygiene education , sanitation ,	0,0000000000000000000000000000000
285	민자주방망이버섯의 폴리페놀 함량과 항산화활	한국식품영양과학회지 = J	v.35,0 no.10,0	이양숙 ; 주은영 ; 김	Lepista nuda , vitamin C , polypl	0,0000000000000000000000000000000
286	가솔린 엔진 장착 차량에서 과도구간 제어특성	한국자동차공학회논문집 =	v.16,0 no.5,0	정연식 ; 박진일 ; 이	과도상태 , 점화시기 , 제동 연료	0,0000000000000000000000000000000
287	험업시 엔진 마찰이 차량 모드 연비에 미치는 영	한국자동차공학회논문집 =	v.16,0 no.6,0	임건병 ; 위효성 ; 브	에너지소비효율 , 시뮬레이션 ,	0,0000000000000000000000000000000
288	추출 조건에 따른 어수리의 항산화 활성 및 성분	한국식품저장유통학회지 =	v.16,0 no.5,0	방지은 ; 최해연 ; 김	anti-oxidative activity , Heracleu	0,0000000000000000000000000000000
289	천연물 유래 항균물질을 포함한 가식성 Pullulan	한국식품영양과학회지 = J	v.38,0 no.10,0	김기명 ; 황권택 ; 유	anticariogenic edible film , grape	0,0000000000000000000000000000000
290	염농도와 절임온도에 따른 동배기의 품질특성	한국식품저장유통학회지 =	v.16,0 no.5,0	김도훈 ; 윤광섭	Dombaegi , salting condition , qu	0,0000000000000000000000000000000
291	저장방법에 따른 '세토카' 감귤의 품질변화	한국식품저장유통학회지 =	v.16,0 no.5,0	이지현 ; 안현주 ; 이	Setoka , citrus , storage	0,0000000000000000000000000000000
292	황금추출물이 고지방 식이를 급여한 흰쥐의 지질	한국식품영양과학회지 = J	v.39,0 no.2,0	윤혜진 ; 박영숙	Scutellaria baicalensis , lipid per	0,0000000000000000000000000000000
293	인천지역 대표 중고등학교 여자 선수의 운동경기	한국식품영양과학회지 = J	v.34,0 no.3,0	성현이 ; 장경자	elite female adolescent athlete ,	0,0000000000000000000000000000000
294	유자의 화학적 특성 및 아질산염 소거작용	한국식품영양과학회지 = J	v.34,0 no.4,0	신정혜 ; 이준열 ; 주	citron (Citrus junos) , electron d	0,0000000000000000000000000000000
295	미량원소 강화 식품소재의 항균효과	한국식품영양과학회지 = J	v.35,0 no.1,0	김보미 ; 목종수 ; 오	antibacterial activity , trace elem	0,0000000000000000000000000000000
296	Pepsin에 의한 Zein 가수분해물의 항균활성	한국식품영양과학회지 = J	v.35,0 no.2,0	강윤정 ; 이상덕 ; 이	zein hydrolysate , pepsin , antib	0,0000000000000000000000000000000
297	대구지역 학교급식 식단에 대한 고등학생의 기호	한국식품영양과학회지 = J	v.35,0 no.7,0	김소희 ; 차명화 ; 김	preferences , food intake , scho	0,0000000000000000000000000000000
298	Cook-chill 파전의 관능성 및 안전성 평가	한국식품영양과학회지 = J	v.34,0 no.5,0	류은순 ; 정동관	cook/chill Pajeon , sensory eva	0,0000000000000000000000000000000
299	머루 과피 용매추출물의 항산화성 및 아질산염	한국식품영양과학회지 = J	v.35,0 no.8,0	최선영 ; 조현소 ; 성	wild grape skin , electron donatir	0,0000000000000000000000000000000
300	보리 도정 겨의 첨가가 쿠키와 머핀의 품질에 미	한국식품영양과학회지 = J	v.33,0 no.8,0	김준희 ; 이영택	barley bran , sugar-snap cookie	0,0000000000000000000000000000000
301	감마선 조사가 감귤 정유의 생리활성에 미치는	한국식품영양과학회지 = J	v.34,0 no.6,0	김현주 ; 조철훈 ; 이	citrus essential oil , physiologica	0,0000000000000000000000000000000
302	Anti-Helicobacter pylori 항체를 함유한 면역우유	한국식품영양과학회지 = J	v.35,0 no.8,0	정은주 ; 박나영 ; 이	Helicobacter pylori , yogurt , imn	0,0000000000000000000000000000000
303	건식절임법으로 제조한 오이지의 절임조건에 따	한국식품영양과학회지 = J	v.34,0 no.5,0	김청희 ; 양윤형 ; 이	pickled cucumber , dry salting m	0,0000000000000000000000000000000
304	감마선 조사에 의한 복분자 착즙액 에탄올 추출	한국식품영양과학회지 = J	v.35,0 no.3,0	김희정 ; 조철훈 ; 김	Rubus coreanus Miq. , biologica	0,0000000000000000000000000000000
305	Cook-chill 부추전의 관능성 및 안전성 평가	한국식품영양과학회지 = J	v.34,0 no.6,0	이경은 ; 류은순 ; 정	cook-chilled Buchu-jeon , sens	0,0000000000000000000000000000000
306	에탄올에 의해 유발된 간독성에 대한 모과 추출	한국식품영양과학회지 = J	v.35,0 no.10,0	이유미 ; 이재준 ; 신	Chaenomeles sinensis Koehne ,	0,0000000000000000000000000000000
307	Corn Bran Fiber를 이용한 저지방 머핀의 품질 특	한국식품영양과학회지 = J	v.34,0 no.5,0	정진영 ; 김선아 ; 정	corn bran fiber , low-fat , muffin	0,0000000000000000000000000000000
308	세척 및 시판 수삼 이용실태와 소비자 인식도 분	한국식품저장유통학회지 =	v.16,0 no.4,0	조순덕 ; 장민선 ; 김	fresh ginseng , consumer , prefe	0,0000000000000000000000000000000
309	상황버섯(Phellinus linteus) 추출물이 고지방,고	한국식품영양과학회지 = J	v.39,0 no.1,0	송원영 ; 성병훈 ; 김	Phellinus linteus , lipid compositi	0,0000000000000000000000000000000
310	가공오징어의 섭취가 쥐의 혈중 지질 조성 및 항	한국식품영양과학회지 = J	v.39,0 no.3,0	정향숙 ; 하지혜 ; 오	cholesterol , cuttlefish , lipid met	0,0000000000000000000000000000000
311	고지방 식이로 유도된 비만쥐에서 식이 베타글루	한국식품영양과학회지 = J	v.31,0 no.6,0	강순아 ; 장기효 ; 홍	obesity , <TEX>$₩₩beta$</TE	0,0000000000000000000000000000000
312	대황유래 Laminaran이 고콜레스테롤 식이를 급여	한국식품영양과학회지 = J	v.35,0 no.7,0	김영명 ; 한찬규 ; 변	Eisenia bicyclis , laminaran , hyp	0,0000000000000000000000000000000
313	홍삼추출잔사로부터 항암면역조절 활성을 보여	한국식품영양과학회지 = J	v.32,0 no.5,0	곽이성 ; 신한재 ; 손	Panax ginseng , red ginseng by-	0,0000000000000000000000000000000
314	케일녹즙 분말식이가 흰쥐의 혈중 지질, 엽산 및	한국식품영양과학회지 = J	v.34,0 no.8,0	정은정 ; 김수연 ; 노	kale juice powder , cholesterol ,	0,0000000000000000000000000000000
315	비만 여대생에 대한 체중조절 프로그램의 적용	한국식품영양과학회지 = J	v.34,0 no.9,0	서지형	weight control , overweight , obe	0,0000000000000000000000000000000
316	무청이 고콜레스테롤 식이 흰쥐 간조직의 항산화	한국식품영양과학회지 = J	v.34,0 no.8,0	이순재 ; 안정모 ; 구	radish loaves powder , antioxida	0,0000000000000000000000000000000
317	야콘 분말이 고지방-고콜레스테롤 식이를 급여	한국식품영양과학회지 = J	v.39,0 no.2,0	김아라 ; 이재준 ; 이	Yacon (Polymnia sonchifolia) ,	0,0000000000000000000000000000000
318	Chromium Methionine첨가가 비만지수 및 지질 대	한국식품영양과학회지 = J	v.32,0 no.5,0	김현숙 ; 이선연 ; 김	chromium methionine , high fat c	0,0000000000000000000000000000000
319	신선편이 양파 가공작업장 내의 시설 및 공정별	한국식품저장유통학회지 =	v.16,0 no.4,0	이혜옥 ; 김지영 ; 듬	fresh-cut onion , microbial conta	0,0000000000000000000000000000000
320	신선편이 채소류 가공작업장 내 시설 및 제품의	한국식품저장유통학회지 =	v.16,0 no.4,0	김병삼 ; 이혜옥 ; 김	fresh-cut leaf vegetables , micro	0,0000000000000000000000000000000
321	영양교육과 운동프로그램이 과체중이상의 성인	한국식품영양과학회지 = J	v.39,0 no.1,0	김은정 ; 황혜진	nutrition education , exercise prc	0,0000000000000000000000000000000
322	산지별 유자의 생리활성	한국식품영양과학회지 = J	v.38,0 no.11,0	신정혜 ; 이수정 ; 김	yuza , electron donating ability ,	0,0000000000000000000000000000000

〈표 8〉 김OO 상위 유사확률(추천대상)

	B	C	E	F	G	J
1	TITLE	JOURNALNAME	VOLUMENAME	AUTHOR	KEYWORDS	김OO 유사확률
2	채소류 메탄올 추출물의 In Vitro와 Cell System에	한국식품영양과학회지 = Jo	v.34.0 no.8.0	이영아 ; 김현영 ; 조	antioxidative activity , vegetable	0.00000003796864701004490000000
3	Cook/Chill System에서의 고기완자 제조에 관한	한국식품영양과학회지 = Jo	v.32.0 no.5.0	강은주 ; 김선영 ; 육	Wanjaiun , cook/chill system , a	0.00000003796864701004490000000
4	참나무 부위별 추출물의 유지에 대한 항산화 효	한국식품영양과학회지 = Jo	v.34.0 no.7.0	우나리아 ; 김태수 ;	Crotalaria sessiflora L , solvent e	0.00000000021947194312936100000
5	키토산-아스코베이트의 용해성, 항산화성 및 항	한국식품영양과학회지 = Jo	v.35.0 no.8.0	이승배 ; 이혜경 ; 김	chitosan-ascorbate , solubility ,	0.00000000021947194312936100000
6	뽕즙이 다른 오디추출물의 Helicobacter pylori에	한국식품영양과학회지 = Jo	v.35.0 no.1.0	조영제 ; 천성숙 ; 0	mulberry fruits , Helicobacter pyl	0.00000000021947194312936100000
7	음증별 치곤 추출물의 생리활성 탐색	한국식품저장유통학회지 = Ko	v.16.0 no.6.0	이광재 ; 박민희 ; A	Chicon , Belgium endive , antio	0.00000000021947194312936100000
8	황기 효소분해물 열수추출액의 이화학적 특성 및	한국식품영양과학회지 = Jo	v.39.0 no.3.0	권상철 ; 최구희 ; 홍	Astragalus membranaceus , enzy	0.00000000021947194312936100000
9	초피나무 추출물의 항산화, 항염증 및 항혈전 효	한국식품영양과학회지 = Jo	v.35.0 no.1.0	장미진 ; 이순재 ; 조	Zanthoxylum piperitum , antioxid	0.00000000021947194312936100000
10	전통적 덕추증자법으로 처리한 홍삼의 일부 항산	한국식품영양과학회지 = Jo	v.34.0 no.5.0	예은주 ; 김수정 ; 부	red ginseng , Processing metho	0.00000000021947194312936100000
11	초임계 이산화탄소를 이용한 겨우살이(Mistletoe	한국식품영양과학회지 = Jo	v.39.0 no.1.0	장태오 ; 유양훈 ; 홍	Viscum album , supercritical carb	0.00000000021947194312936100000
12	5종의 항신료 메탄올 추출물의 항산화 효과	한국식품영양과학회지 = Jo	v.33.0 no.9.0	김진 ; 김성애 ; 윤혼	antioxidative activity , DPPH rad	0.00000000021947194312936100000
13	항신료 추출물의 항산화활성 및 Helicobacter py	한국식품영양과학회지 = Jo	v.35.0 no.3.0	차원섭 ; 김정환 ; 0	antimicrobial , antioxidative activ	0.00000000021947194312936100000
14	큰잎모자반 효소적 추출물의 항산화 활성	한국식품영양과학회지 = Jo	v.39.0 no.4.0	고석천 ; 강성명 ; 은	Sargassum coreanum , antioxida	0.00000000021947194312936100000
15	다양한 용매에 의해 추출된 오미자의 항산화능고	한국식품영양과학회지 = Jo	v.39.0 no.4.0	김민정 ; 박은주	Omlia (Schizandra chinensis B.)	0.00000000021947194312936100000
16	천연 식물자원으로부터 Acetylcholinesterase 저	한국식품영양과학회지 = Jo	v.34.0 no.3.0	김대익 ; 이성현 ; 추	plant extract , anti dementia , an	0.00000000021947194312936100000
17	수확시기가 다른 품종별 구기자 추출물의 항산화	한국식품영양과학회지 = Jo	v.35.0 no.9.0	박수진 ; 박원종 ; 0	antioxidant , Lycium chinensis M	0.00000000021947194312936100000
18	녹차첨가가 청국장의 관능적 품질 개선에 미치는	한국식품영양과학회지 = Jo	v.35.0 no.4.0	김재훈 ; 김선임 ; 김	Chungkookjang , green tea , ser	0.00000000021947194312936100000
19	주름 미더덕(Styela plicata) 추출물의 항산화력	한국식품영양과학회지 = Jo	v.34.0 no.7.0	김진주 ; 김선정 ; 김	Styela plicata , antioxidant , anti	0.00000000021947194312936100000
20	시판 쌀 메탄올 추출물의 항혈전 및 항산화 활성	한국식품영양과학회지 = Jo	v.34.0 no.5.0	손호용 ; 권경숙 ; 윤	antithrombosis , antioxidant , me	0.00000000021947194312936100000
21	구기자(Lycii fructus) 추출물의 항산화와 항고혈	한국식품영양과학회지 = Jo	v.34.0 no.9.0	조영제 ; 천성숙 ; 조	Lycii fructus , physiological activ	0.00000000021947194312936100000
22	녹차의 음용을 위한 최적 침출조건 예측	한국식품영양과학회지 = Jo	v.35.0 no.6.0	장문조 ; 하현절 ; 윤	green tea , leaching condition ,	0.00000000021947194312936100000

협업필터링을 통해 학습한 논문이 다르고 이를 통해 추출한 자질도 상이하기 때문에 발생하는 당연한 결과다.

5.3 추천논문 만족도 조사결과

본 모델을 이용하여 실험대상자 5명 각각에 대하여 상위 20편의 추천논문을 생성한 다음, 추천논문 리스트를 이메일로 발송한 후 5점 척도(매우만족 5, 만족 4, 보통 3, 불만족 2, 매우 불만족 1)로 만족도를 조사한 결과 〈표 9〉와 같이 상당히 만족(4.75)하는 것으로 조사되었다. 특히, 식품분야 연구자들이 전기 분야 연구자들보다 만족도가 더 높게 나타났다. 이것은 실험대상자에 포함된 식품분야 연구자들이 전기 분야 연구자들보다 더 많은 논문을 저술하고 NDSL을 많이 이용함으로써, 협업필터링 방식에 의거 더 많은 논문 상호추천과 논문을 학습한 결과다.

〈표 9〉 추천논문 만족도 조사결과

대상자	연구분야	만족도(5분 척도)	첨언
강OO	식품	매우만족(5)	개인맞춤형 자동추천서비스는 매우 좋은 일이라고 생각하며, 추천 논문에 매우 만족함.
김OO	식품	매우만족(5)	별도의 검색 없이 관련 논문을 볼 수 있으면 아주 좋을 것 같음. 최근 5년으로 추천논문을 제한하면 더 좋겠음.
김OO	전기, 전자	만족(4)	연구 분야가 전기, 전자인데 생소한 식품관련 논문도 일부 있음
서OO	기계	매우만족(5)	추천결과에 매우 만족함.
황OO	전기	무응답	-

6. 결론

이용자들이 원하는 정보를 예측하여 관심 가질만한 정보를 자동으로 추천하는 시스템은 전문가시스템, 데이터마이닝, 정보검색 등 다양한 분야에서 오래 전부터 연구하여 왔다. 최근에는 콘텐츠기반추천시스템과 협업필터링을 결합하거나 다른 분야 모델을 접목한 하이브리드 추천시스템 연구에 치중하고 있다.

본 논문은 기존 추천시스템 문제점을 해결하고 대규모 정보센터나 도서관에서 효율적, 지능적으로 학술논문을 추천할 수 있도록 협업필터링과 나이브베이즈모델을 결합한 새로운 방식의 추천모델을 제시하였다. 본 방식은 관심사항이나 프로파일을 별도로 등록·관리하지 않더라도 협업필터링을 통해 동료 연구자들이 저술하거나 열람한 논문을 자동으로 추천 및 학습할 수 있다. 아울러 기존 콘텐츠기반추천시스템의 과도한 특성화(Over-specialization) 문제와 이용로그나 평가정보가 축적되기 전까지 신규논문을 추천할 수 없었던 협업필터링 방식의 문제점을 동시에 해소하였다.

본 모델을 한국과학기술정보연구원 'NDSL 학술논문'에 적용하여 실험하였다. 실험은 식품과 전기 분야 11개 국내 학술지를 대상으로 하였으며 현재 NDSL을 이용하는 해당분야 연구자 4명에게 피드백을 받은 결과 추천논문에 대체로 만족(4.75)하는 것으로 나타났다. 향후 시스템 성능을 보다 객관화하기 위해 실험대상을 확장하고, 다양한 분야에서 활동하는 연구자들을 활용하여 추천모델 성능을 개선하는 추가적인 연구가 필요하다.

참고문헌

김명찬. 2003.『텍스트 마이닝을 이용한 스팸메일 방지 기술』. 정보통신부.

김성희, 엄재은. 2008. 기계학습을 이용한 문서 자동분류에 관한 연구,『정보관리연구』. 39(4): 47-66.

김재경, 안도현, 조윤호. 2003. 인터넷 쇼핑몰을 위한 데이터마이닝 기반 개인별 상품추천 방법론의 개발.『한국지능정보시스템학회논문지』, 9(3): 177-191.

김찬주, 황규백. 2009. 소셜 북마킹 시스템의 스패머 탐지를 위한 기계학습 기술의 성능 비교.『정보과학회논문지: 컴퓨팅의 실제 및 레터』, 15(1): 345-349.

김판준, 이재윤. 2007.『문헌간 유사도를 이용한 자동분류에서 미분류 문헌의 활용에 관한 연구. 정보관리학회지』, 24(1): 251-252.

박지선, 김택헌, 류영석, 양성봉. 2002. 추천시스템을 위한 2-way 협동적 필터링 방법을 이용한 예측알고리즘,『정보과학회지』, 29(9): 669-675.

안신현. 2007. 콘텐츠 추천 시스템에 관한 이슈. [cited 2010. 6. 10]. 〈http://beatlover.kaist.ac.kr/resource/ash_memo_05.pdf〉

연종흠, 심준호, 이상구. 2010. 확장된 나이브

베이즈 분류기를 활용한 질문-답변 커뮤니티의 질문 분류.『정보과학회논문지: 컴퓨팅의 실제 및 레터』, 16(1): 95-100.

이재문. 2003. 휴리스틱을 이용한 KNN의 효율성 개선.『정보처리학회논문지』, 10-B (6): 719-720.

이충무, 이상기, 이병섭. 2009. 추천시스템과 소셜 네트워크를 융합한 지능형 연구자연결망 구축.『정보관리연구』, 40(1) : 199-215.

정영미, 이용구. 2005. 정보검색 성능 향상을 위한 단어 의미 중의성 해소 모형에 관한 연구.『정보관리연구』, 22(2): 125-145.

최종후. 2000.『AnswerTree를 이용한 데이터 마이닝 의사결정나무분석』. 서울:SPSS 아카데미.

한성희. 2009.『추천시스템기술소개』. KBS 방송기술연구소.

홍태호. 2005. 클러스터링 기반 사례기반추론을 이용한 웹 개인화 추천시스템.『한국지능정보시스템학회논문지』, 11(1): 107-121.

Balabanovic, M., and Shoham Y. 1997. "Fab : Content-Based, Collaborative Recommendation." *Communications of the Association for Computing Machinery,* 40(3): 66-72.

Baumann, S., and Hummel, O. 2005. "Enhancing Music Recommendation Algorithms Using Cultural Metadata." *Journal of New Music Research,* 34(2): 161-172.

Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R, and J.H. Suh. 2002. "A Personalized Recommendation Procedure for Internet Shopping Support." *Electronic Commerce Research and Applications,* 1(3): 301-313.

Konstan, J.A., Miller, B.N., Maltz, D. 1997. Herlocker, J.L., Gordon, L.R., and Riedl, J. 1997. "GroupLens: Applying Collaborative Filtering to Usenet News." *Communications of the Association for Computing Machinery.* 40(3): 77-87.

Krulwich, B., and Burkey, C. 1996. "Learning user information interests through extraction of semantically significant phrases." *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access.*

Li, Q., and B.M. Kim. 2003. "Clustering Approach for Hybrid Recommender System." *Proceedings of IEEE/WIC International Conference on Web Intelligence,* 33-39.

Miyahara, K., Pazzani, M.J. 2000. "Improvement of Collaborative Filtering with the Simple Bayesian Classifier". *Lecture notes in computer science,*

1886.

Pedro Domingos and Michael Pazzani. 1996. "Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier." in *Proceedings of the 13th International Conference on Machine Learning.* 105-122.

Robles, V., Larranaga, P., Menasalvas, E., Perez, M.S., Herves, V. 2003. "Improvement of Naive Bayes Collaborative Filtering Using Interval Estimation." *Proceedings of the 2003 IEEE/WIC International Conference on Web Intelligence,* 168-174.

Roh, T.H., Oh, K.J., and Han, I., 2003. "The Collaborative Filtering Recommendation Based on SOM Cluster-Indexing CBR." *Expert Systems with Applications,* 25(3): 413-423.

Sarwar, B.M., Karypis, G., Konstan, J.A., and Riedl, J., 2001. "Item-Based Collaborative Filtering Recommendation Algorithms." *Proceedings of the 10th International World Wide Web Conference,* 285-295.

Schafer, J.B., Konstan, J.A., Riedl, J., 1999. "Recommender Systems in e-Commerce." *Proceedings of the ACM Conference on Electronic Commerce.*

Shardanand, U., and Maes, P. 1995. "Social Information Filtering: Algorithms for Automating." *Proceedings of the SIGCHI conference on Human factors in computing systems.* 210-217.

Sheth, B., and Maes, P. 1993. "Evolving agents for personalized information filtering." *Proceedings of the 9th IEEE Conference on Artificial Intelligence for Applications.*

Weng, S.-S., and M.-J. Liu. 2004. "Feature-Based Recommendations for One-to-One Marketing." *Expert Systems with Applications,* 26(4): 493-508.