

# 선호도 추정모형과 협업 필터링기법을 이용한 고객추천시스템

신태수  
연세대학교 정경대학 경영학부  
(tsshin@yonsei.ac.kr)

장근녕  
연세대학교 정경대학 경영학부  
(knchang@yonsei.ac.kr)

박유진  
연세대학교 정경대학 경영학부  
(parkyoujin@yonsei.ac.kr)

본 연구는 상품추천을 위해 필요한 고객 선호도 추정모형(Customer Preference Estimation Model)을 제안하고, 이러한 선호도 추정결과에 따른 선호도 정보를 이용하여 궁극적으로 상품추천의 성과를 제고시키기 위한 방법을 제시하였다. 즉, 제품에 대한 고객 선호 영향요인들과 고객 선호도와와의 관계를 모형화 함으로써 고객 선호도를 보다 더 정확히 추정할 수 있는 새로운 선호도 추정모형을 제안하였다. 이 제안모형은 선호도 영향요인들의 상대적인 가중치를 선호도 최적화 학습을 통해 도출함으로써, 보다 정확한 선호도 측정을 가능하게 해 준다. 한편, 이 모형의 타당성을 검증하기 위해서 본 연구에서는 가상서점 고객들을 대상으로 고객 선호도 정보를 수집한 후, 본 제안모형을 적용했을 때의 협업 필터링의 추천성과 사전가중치 부여방식인 기존 선호도 계산식을 이용했을 경우의 추천성과를 비교 분석하였다. 이에 대한 실증분석 결과는 본 연구에서 제안한 선호도 추정모형을 적용했을 때의 협업 필터링의 성과가 기존 선호도 계산방식을 적용했을 때의 협업 필터링의 성과보다 더 우수한 것으로 나타났다.

논문접수일 : 2005년 11월      게재확정일 : 2006년 12월      교신저자 : 신태수

## 1. 서론

현재 고객추천시스템에서 대표적으로 적용되고 있는 협업 필터링기법(collaborative filtering technique)은 사용자의 선호도에 대한 데이터를 기반으로 사용자가 관심을 가질 것으로 생각되는 아이템을 추천해 주는 기법으로서 상품추천을 위한 개인화(personalization)기법의 대표적인 유형이다.

그러나, 협업 필터링을 적용할 경우에 발생하는 문제점 중의 하나는 고객들이 개인화 서비스를 받기 위해서는 일단 등록시에 다수의 상품에 대한 선호도 평가점수를 입력해야 하는 수고가

필요하다는 것이다. 이러한 문제점을 피하기 위한 방안의 하나로는 고객에게 등록시 선호도 평가점수를 입력 받지 않고, 구매 데이터를 사용하여, 이를 상품 선호도에 대한 이진 평가점수로 생각하여 고객간 상관계수 계산을 수행하는 방안이 있다(Breese et al., 1998, 김종우 등, 2004). 그러나, 이처럼 고객의 평가점수를 사용하지 않고, 구매 이력을 사용하는 추천방법은 구매 이력이 없는 신규 고객에게는 적용할 수 없다. 또한, 신규고객이 아닌 경우라도, 최근 몇 년 동안 거래가 없었으나, 사이트 방문 등을 통해 계속 관심품목을 갖고 있는 고객의 경우에도 구매 이력만을 사용하는 추천방법을 적용하기가 어렵

다. 그 밖의 방법으로는, 고객의 선호도를 나타내는 다양한 정보를 하나의 지표로 반영함으로써 이를 선호도로서 추정하는 방법이 있다.

일반적으로 고객 선호도란 고객이 제품에 대해 직접 평가한 점수로서 정의되고 있으며, 이러한 평가점수는 개인의 주관적인 선호도로 파악된다. 그러나, 고객의 직접적인 선호도 측정이 아닌, 인터넷상의 고객의 행동을 파악함으로써, 고객의 제품에 대한 선호성향을 보다 객관적인 자료를 통해 추정할 수 있다. 예를 들어, 개인의 각 제품에 대한 선호도를 직접 개인들로부터 물어 보는 것이 아니라, 개인의 초기 프로파일로부터 얻을 수 있는 관심제품군에 대한 정보, 과거의 구매정보, 상점방문횟수, 제품정보 조회횟수 등을 분석함으로써 고객의 제품에 대한 선호도를 추정할 수 있다(Kuo and Chen, 2001; 김종우 등, 1999).

고객 선호도를 선호도에 영향을 주는 요인들을 토대로 추정할 경우, 고객 선호도는 이들 요인들의 단순한 합이 아니라, 요인들간의 상대적인 중요도를 정확히 고려하여 계산되어야 한다. 그러나, 기존연구에서는 이러한 상대적인 가중치를 전문가의 판단에 의존하거나 고정된 상수 값 등으로 정의함으로써(김종우 등, 1999), 이러한 가중치 정보가 고객추천시스템의 성과를 향상시킬 수 있는 중요한 의사결정 변수로서는 거의 고려되지 않았다. 이처럼, 고객의 선호도 계산에 반영되는 다양한 고객 정보들의 상대적인 강도를 나타내는 가중치를 고객의 개인적인 상황이나 주위 환경 변화에 따라 조정하지 않고 항상 고정된 값으로 사용할 경우, 이러한 선호도 계산방법을 통해 얻어진 선호도는 지속적으로 변화하는 고객의 관심을 충분히 반영하지

못하는 문제가 존재한다(장근녕, 박유진, 2002).

본 연구는 고객의 선호도 영향요인들간의 상대적 가중치를 합리적으로 도출하고, 이를 통해 고객 선호도를 정확히 추정하기 위한 방안으로서, 새로운 고객 선호도 추정모형을 제안한다. 이 모형은 특히 고객이 직접 평가한 제품에 대한 선호도와 선호도 영향요인들간의 인과관계를 모형화함으로써, 선호도 영향요인들간의 상대적인 가중치를 추정하는 기능을 수행한다. 따라서, 본 연구에서 제안한 선호도 추정모형은 기존의 선호도 추정방식의 문제점을 보완함으로써, 궁극적으로 고객추천시스템의 성과를 제고시킬 수 있다. 이를 입증하기 위해서 본 연구는 가상서점 고객들을 대상으로 자료를 수집하여, 본 연구모형을 적용했을 때의 실증분석 결과를 제시하고, 사전 가중치부여 방식의 고객 선호도 계산식을 이용한 기존 추천시스템방식의 추천성과와의 비교분석을 통해 본 연구모형의 타당성을 검증한다.

본 논문의 구성은 다음과 같다. 제 2절에서는 고객추천시스템에 대한 선행연구들을 고찰하고, 제 3절에서는 연구모형을 제시한다. 제 4절은 실증분석 및 결과를 요약하고, 제 5절에서는 결론을 제시한다.

## 2. 고객추천시스템에 관한 연구

개인화된 정보의 추천은 고객과의 상호작용에 있어 가장 중요한 비즈니스 활동이다. 웹의 출현과 함께 인터넷 업체들은 훨씬 더 낮은 원가를 투입하여 고객이 어떠한 분야를 더욱 선호하는지에 관한 정보를 얻을 수 있게 되었으며

이러한 정보를 바탕으로 실시간으로 일대일(one-to-one) 원리에 따라 정보를 추천할 수 있게 되었다(Cheung et al., 2000). 보다 넓은 의미에서 이와 같은 추천 기술들은 개별 고객에게 가장 적합한 웹 페이지의 제공을 위해 협조하기 때문에 웹사이트 개인화의 한 부분이라 할 수 있다. 보편적으로 이러한 시스템을 추천시스템(recommender system)이라 하고 Amazon.com, CDNOW, Drugstore.com, eBay, MovieFinder.com, Reel.com 등과 같은 인터넷 업체들은 그들의 웹사이트에 하나 또는 그 이상의 다양한 추천시스템 기술을 사용하고 있다(Schafer et al., 2001).

개인화된 정보를 추천하는 대표적인 기법으로는 내용기반 필터링기법(content based filtering technique), 협업 필터링기법(collaborative filtering technique), 규칙기반 필터링기법(rules based filtering technique), 혼합형 필터링기법(hybrid filtering technique) 등을 들 수 있다(Kim et al., 2001; Mooney and Roy, 2000; Pazzani, 1999; Pazzani et al., 1996; 김종우 등, 1999). 이들 기법에 대한 기능 및 특징은 다음과 같다.

## 2.1 제품, 서비스 추천을 위한 단일 필터링기법

내용기반 필터링기법은 추천하고자 하는 상품 자체의 설명과 사용자 프로파일간의 유사성을 고려하여 추천하는 방식으로서 제품의 속성과 고객 관심의 매칭을 통해 정보를 추천하며 문서 내 단어의 중요도에 의해 문서를 표현한다. Fab(Balabanovic, 1997; Balabanovic and Shoham, 1997)은 가장 높은 TF-IDF(Salton and McGill, 1983) 가중치를 기록한 100개의 단어로 문서를

표현하였으며, Syskill과 Webert(Pazzani et al., 1996)는 128개의 가장 유의한 단어들에 의해 문서를 표현하였다. 그러나 내용기반 필터링은 문서 내에 존재하는 단어의 특징 추출을 통해 문서를 분류하고 추천하기 때문에 문서 내에 너무 많은 단어들 존재한다면 문서의 분류 작업이 어려워진다. 뿐만 아니라 특징이 너무 많을 경우에는 유용한 정보 선택을 위한 사전 준비단계를 필요로 하며 반대로 특징이 너무 적을 경우에는 고객의 프로파일 학습이 어려워진다.

내용기반 필터링기법은 사용자의 요구사항이나 성향(기호) 등의 정보를 포함하고 있는 사용자 프로파일을 필요로 한다. 시스템은 사용자가 관심을 보였던 정보나 상품의 내용을 분석하여 성향을 학습함으로써 사용자 프로파일을 구축하고, 해당 프로파일과 유사한 정보나 상품만을 추천한다. 전자우편이나 유즈넷 뉴스 정보를 여과하기 위한 에이전트나 Lieberman의 Letizia는 사용자가 과거부터 가져왔던 웹문서에 대한 관심 정도에 기반하여 문서를 분류, 삭제, 또는 추천하기 위해 학습기법을 사용했다(Liebermann, 1997; Maes, 1994). 유사한 방법으로 Cohen의 Ripper 시스템도 전자우편을 분류하기 위해 유사한 방법을 채택했다(Cohen, 1996).

이상의 내용기반 필터링기법은 추천시스템에서 중요한 역할을 해왔지만 정보나 상품에 대한 내용정보를 필요로 하고, 사용자 중심적인 추천만이 이루어지기 때문에 추천의 폭이 제한적일 수 밖에 없는 단점이 있다.

반면에, 협업 필터링기법은 해당 고객과 선호도가 유사한 고객들의 상품에 대한 평가점수를 활용하여 고객에게 적합한 상품정보를 제공한다. 즉, 기호가 유사한 사용자의 프로파일을 비

교하여 추천하는 방식으로서 정보나 상품에 대한 사용자의 의견을 데이터베이스화 한다. 이 데이터베이스를 이용하여 특정 정보나 상품에 대한 의견이 유사한 사용자를 발견하고, 유사한 성향을 가지고 있는 사용자들이 접근 또는 구입한 정보나 상품을 조사하여 교차 추천하는 방식이다. 즉, 고객 사이에서 발견된 상관관계를 이용하여 비슷한 취향을 가진 고객들간의 비교를 통해 새로운 고객이 흥미를 가지거나 구매할 가능성이 높은 정보를 예측하고 추천한다(Pazzani, 1999).

초기의 협업 필터링 시스템인 While Tapestry는 사용자가 명시적으로 정보에 대한 평가를 하였으나, GroupLens는 사용자의 명시적 평가 없이 예측을 수행하였다. Ringo와 Bellcore의 Video Recommender는 각각 음악과 영화를 대상으로 추천서비스를 제공한 시스템이다(Goldberg et al., 1992; Herlocker et al., 1999; Hill et al., 1995). 그 밖에 Lotus Notes, Firefly, SiteSeer, PHOAKS, Grassroots 등과 같은 다양한 추천시스템들에서 협업 필터링이 사용되고 있다.

이러한 방법들은 유사한 성향을 지닌 다른 사용자에게 의해 추천을 받는 형식이므로 내용기반 필터링기법보다는 다양한 추천을 받을 수 있고, 추천되는 정보의 내용을 고려할 필요가 없다는 장점이 있다. 그러나, 새로운 정보가 생성되었을 경우 누군가에 의해 접근되고 평가되기까지는 아무에게도 추천될 수 없는 문제점(first-rater problems)을 가지고 있다(Balabanovi and Shoham, 1997; Good et al., 1999; Sarwar et al., 1998). 그리고, 특이한 성향의 사용자는 유사한 사용자를 찾을 수 없으므로, 추천의 대상에서 제외될 수 있다는 단점이 있다(Balabanovi and Shoham, 1997). 또한, 협업 필터링의 자료의 희소성(sparsity)문제

로 인해서 새로운 정보가 여러 사용자에게 의해 평가되기 전까지는 인기 있는 주제만이 추천되거나 혹은 동일한 웹 페이지가 계속해서 추천될 수도 있다(Mooney and Roy, 2000).

그 밖의 고객 추천시스템을 위한 기법으로는 규칙기반 필터링기법이 있으며, 이 방법은 사용자의 프로파일이나 행위에 근거한 조건문 형식의 규칙을 이용하여 개인화된 추천을 제공하는 방법으로 가장 일반적이고 전통적인 방법이다. Broad-Vision의 One-to-One 시스템은 규칙기반 필터링을 사용하여 고객에게 적합한 개인화된 광고를 추천하고 있다(BroadVision, 1996). 규칙기반 필터링에서 마케팅 전문가로부터 도출된 마케팅 규칙은 개인화된 추천을 위한 가장 핵심 요소이며 그 유효성은 규칙베이스 내의 지식의 질(quality)에 달려있다. 그러나 마케팅 전문가로부터 가치 있는 마케팅 규칙의 도출, 도출된 규칙의 유효성 평가, 지식베이스 내의 마케팅 규칙의 지속적인 갱신 및 변환과 같은 일들은 상당히 어려운 작업으로 이는 규칙기반 필터링의 문제점으로 지적될 수 있다(Kim et al., 2001).

최근에는 규칙기반 필터링기법에서 규칙을 생성시키는 방법으로서 데이터마이닝 기법을 이용하여 규칙을 자동 생성시키는 방법이 활용되고 있다. 여기에는 의사결정트리(Cho et al., 2002; Kim et al., 2001), 연관성 규칙 분석기법(Changchien and Lu, 2001; Lee et al., 2001; Schafer et al., 1999; Wang and Shao, 2004; 김종우, 이경미, 2000), 러프집합분석(Changchien and Lu, 2001) 등이 있다. 예를 들어, Amazon.com에서는 구매자들이 공통적으로 구매한 제품에 대해 연관성 규칙을 적용하여 생성한 추천 상품정보를 제공하는 방식을 적용하고 있다(Schafer et al., 1999).

## 2.2 혼합형(Hybrid) 필터링기법

한편, 기존의 협업 필터링기법과 내용기반 필터링기법에서 생기는 단점을 극복하기 위해 두 가지 방법을 혼합하려는 시도가 있었다(Balabanovi and Shoham, 1997; 이치훈 등, 1999, 정경용 등, 2003). 예를 들어, Fab시스템은 내용기반 필터링기법을 이용하여 웹문서에 대한 관심도를 나타내는 사용자 프로파일을 유지하고, 이를 이용하여 유사한 사용자를 분류한 후 협업 필터링기법을 적용하는 방법을 채택하였다(Balabanovi and Shoham, 1997). 이러한 추천기법은 혼합형 필터링기법으로 분류되어진다. 이처럼, 협업 필터링 자체만으로 이용자의 관심이나 취향에 맞는 아이템에 대한 예측은 잘 수행할 수 있으나 특정 아이템의 내용에 대한 이용자의 정보요구를 해

결하는 것은 효율적이지 못한 문제점을 내용기반 필터링기법과의 결합을 통해 해결할 수 있다. Lee et al.(2001)은 내용기반 필터링기법과 협업 필터링기법의 문제점을 해결하기 위해 고객간의 유사성뿐 아니라, 제품간의 연관성을 고려한 협업 필터링기법과 연관성분석기법을 결합한 추천방식을 제시하였다. 그 밖에, 또 다른 혼합형 추천방식으로서, 사용자기반 협업 필터링기법과 아이템기반 협업 필터링기법을 결합한 추천방식을 통해 두 기법이 가지는 문제점을 해결하기 위한 방법 등이 제시되고 있다(박지선 등, 2002).

위에서 언급한 고객추천기법들을 요약 정리하면, <표 1>에서 보는 바와 같다. 각 고객추천기법에 대한 더 자세한 내용은 관련 문헌을 참조하기 바란다.

<표 1> 고객추천기법의 분류

고객추천기법유형	내용	관련 문헌
내용기반 필터링기법	추천하고자 하는 상품 자체의 설명과 사용자 프로파일 간의 유사성을 고려하여 추천하는 기법	Balabanovic(1997), Balabanovic and Shoham(1997), Cohen(1996), Liebermann(1997), Maes(1994), Pazzani et al.(1996), Salton and McGill(1983)
협업 필터링기법	해당고객과 선호도가 유사한 고객들의 상품에 대한 평가점수를 활용하여 고객에게 적합한 상품정보를 추천하는 기법	Balabanovi and Shoham(1997), Breese et al.(1998), Goldberg et al.(1992), Good et al.(1999), Herlocker et al.(1999), Hill et al.(1995), Mooney and Roy(2000), Pazzani(1999), Sarwar et al.(1998), Shardanand and Maes(1995)
규칙기반 필터링기법	의사결정트리, 연관성 규칙, 러프집합분석 등의 데이터 마이닝 기법을 이용한 고객 추천기법	BroadVision(1996), Changchien and Lu(2001), Cho et al.(2002), Kim et al.(2001), Lee et al.(2001), Schafer et al.(1999), Wang and Shao(2004), 김종우, 이경미(2000)
혼합형 필터링기법	기존의 내용기반 필터링기법과 협업 필터링기법에서 생기는 단점을 극복하기 위한 결합기법 예) 내용기반 필터링기법과 협업 필터링기법을 결합한 기법, 협업 필터링과 연관성분석기법을 결합한 기법, 사용자기반 및 아이템 기반을 결합한 협업 필터링 추천방식 등	Balabanovi and Shoham(1997), Lee et al.(2001), 박지선 등(2002), 이치훈 등(1999), 정경용 등(2003)

본 연구에서는 이상의 추천기법들 중에서 협업 필터링기법을 이용한 추천시스템을 중심으로 본 연구모형을 적용한다. 이를 통해 본 연구모형에서 제안하고 있는 선호도 추정모형을 사용함으로써, 협업 필터링에 의한 추천성과를 얼마나 제고시킬 수 있는지 실증적으로 제시한다.

### 3. 연구모형

본 연구는 제품에 대한 고객 선호도와 제품 선호 영향요인들간에는 인과관계가 존재한다는 점을 고려하여, 미리 수집된 고객 선호도 정보와 제품 선호 영향요인들간의 관계를 모형화한 고객 선호도 추정모형을 제안한다. 기존 연구에서는 단순히 고객의 프로파일, 구매횟수, 웹상의 행동패턴 등과 관련된 정보를 미리 정해진 가중치로 가중평균한 값을 고객 선호도로써 측정하였고, 이때 사용되는 제품 선호 영향요인들간의 상대적 가중치부여에 대한 기준 또한 명확히 제시되지 않았다. 예를 들어, 고객 선호 영향요인에 대한 중요도 분석에 있어서 이들 영향요인들에 대한 가중치를 미리 정해진 기준이나 단순히 동일한 비중으로 선호도를 계산하고 있다. 여기서, 가중치는 고객 프로파일, 구매횟수, 관심표명행위 등이 선호도에 미치는 상대적인 강도를 의미한다. 일반적으로 특정 제품군에 속한 제품의 구매는 해당 제품군의 제품 페이지 방문 행위에 비해서 선호도 점수에 더 많은 영향을 준다고 볼 수 있다. 그러나, 이러한 판단에 대한 실증적 분석이 없이는 구매정보가 과연 다른 행동패턴이나 고객 프로파일 정보와 비교할 때 얼마만큼 고객 선호도를 평가하는데 영향을 줄 것인가를 판단하기 어렵게 된다. 이로 인해, 상품추천을

위한 협업 필터링 등의 분석 등에도 잘못된 결과를 초래할 것이다.

본 연구에서 제안하는 고객추천시스템의 전체적인 구조는 [그림 1]과 같다. 고객추천시스템의 1단계에서는 고객에 대한 제품 선호도 설문 조사를 통해 수집된 고객의 실제 제품 선호도 점수와 각 고객의 프로파일과 구매횟수, 고객 행동패턴정보 등을 이용하여 고객의 제품 선호도 추정모형을 개발한다. 여기서 제품 선호도 추정모형은 다중회귀분석을 통해 구축된다. 2단계에서는 이러한 선호도 추정모형에서 도출된 고객의 선호도에 영향을 주는 요인들에 대하여 통계적 유의수준을 평가한 후, 이중 통계적으로 유의적인 고객 선호 영향요인들에 대해서 각 요인의 가중치를 도출한다. 3단계에서는 추정된 고객 선호 영향 변수들의 상대적 가중치를 이용하여 고객 선호도를 계산한다. 마지막 단계에서는 추정모형을 통해 생성된 선호도 추정치를 협업 필터링에 적용한 후, 이를 이용해 고객에게 상품추천을 수행한다.



[그림 1] 고객추천시스템의 구조

본 연구의 고객추천시스템의 성과를 분석하기 위해서 먼저 가상 인터넷 서점을 구축한 후, 무작위로 추출된 표본 집단을 대상으로 실증분석을 수행하였다. 가상 인터넷 서점은 Windows Server 2000, MS-SQL Server 2000, IIS 5.0, 프론트페이지와 Namo 웹 에디터를 사용하여 구현되었다.

### 3.1 선호도 추정모형

일반적으로 제품에 대한 고객 선호도는 사용자의 선호도를 습득하는 방법에 따라 사용자에게 직접적으로 선호도를 요구하여 습득하는 명시적 선호도(explicit preference rates)와 사용자의 행동을 관찰하여 이를 학습하여 습득하는 암시적 선호도(implicit preference rates)로 나눌 수 있다(Breese et al., 1998; 김진 등, 2000). 예를 들어 상품이나 콘텐츠에 대한 개인별 선호도를 의미하는 고객 프로파일은 고객으로부터 명시적인 방법에 의해 수집하거나 사용자 별로 사이트 접속 패턴을 이용하여 묵시적인 방법으로 수집할 수 있다(Oard and Kim, 1998).

그동안 고객 선호도의 측정방법에 대한 연구에서는 특히 이러한 암시적인 선호도를 측정하는 방법으로서 사용자의 사이트 접속 패턴 등 고객의 웹사이트상에서의 행동패턴에 대한 많은 연구가 제시되어 왔다(Breese et al., 1998; Cho et al., 2002; Lee et al., 2001; 김진 등, 2000).

본 연구에서 분석되는 고객 선호도의 영향요인은 고객 프로파일, 구매횟수, 관심표명행위 등으로 정의되며, 이 영향요인들간의 가중치는 고객 선호도에 미치는 상대적인 강도를 의미하게 된다. 일반적으로 특정 제품군에 속한 제품의 구매는 해당 제품군의 제품 페이지 방문 행위에

비해서 선호도 점수에 더 많은 영향을 준다고 볼 수 있다. 그러나, 얼마만큼 더 영향을 주는지에 대한 분석이 없이는 정확한 고객 선호도에 대한 분석이 어렵게 된다.

한편, 고객의 관심표명행위는 위에서 제시한 웹사이트상에서 이루어진 고객의 행동패턴을 말하며, 이러한 행동패턴은 암시적으로 해당 제품에 대한 선호도 또는 관심도를 보여주는 특성을 갖는다. 이러한 제품에 대한 관심 및 선호도와 관련이 깊은 행동패턴으로는 사이트 클릭횟수, 사이트 체류시간, 장바구니 이용횟수 등을 예로 들 수 있다(Cho et al., 2002; Kuo and Chen, 2001; Lee et al., 2001).

본 연구에서는 고객의 선호도 영향요인들을 고려할 때 필요한 상대적 가중치 결정문제를 선호도 추정모형을 통해 해결한다. 이를 위해서 우선 가상 고객들을 대상으로 관심 제품군에 대한 선호도를 직접 설문조사를 통해 제품들에 대한 실제 선호도를 5점 척도로 측정된 후, 일정기간 동안 고객들로 하여금, 가상 서점 홈페이지에 접속하여 제품구매를 위한 방문을 하도록 하여 해당 웹페이지를 방문하는 동안 발생된 웹정보를 수집하였다. 이들 웹정보는 본 연구에서 고객 선호도를 추정하기 위해 사용되는 영향 변수들로서 사용되며, 이 변수들은 위에서 제시한 고객 프로파일 정보인 관심유무, 해당 서적에 대한 구매횟수, 그리고 고객의 관심표명행위로서 사이트 체류시간, 관심 서적에 대한 클릭횟수, 장바구니 이용횟수 등으로 정의된다. 이러한 변수들에 대한 선정기준은 선행연구(Cho et al., 2002; Lee et al., 2001)에서 제시된 지표들을 기초로 선정되었다. 따라서, 이러한 정보들을 이용하여 추정되는 고객 선호도 추정모형은 아래와 같이 선형 식 (1)로서 유도된다.

$$y = a_1 + b_1 \times x_1 + b_2 \times x_2 + b_3 \times x_3 + b_4 \times x_4 + b_5 \times x_5 \quad (1)$$

- $y$  : 고객의 제품 선호도(1~5)
- $x_1$  : 관심분야에 대한 관심유무(0 또는 1)
- $x_2$  : 사이트 체류시간
- $x_3$  : 관심서적 클릭횟수
- $x_4$  : 장바구니 이용횟수
- $x_5$  : 구매 횟수

식 (1)은 제품에 대한 고객의 선호도와 이에 영향을 주는 변수들간에 선형관계가 존재한다는 가정하에서 유도되는 고객의 제품 선호도에 대한 추정 모형식으로 정의된다. 본 연구에서는 다중회귀모형을 이용하여 이러한 선호도 추정 모형을 개발한다.

### 3.2 협업 필터링기법

일반적으로 협업 필터링은 메모리기반(Memory-based) 필터링에 해당하는 사용자기반 필터링 방식과 모델기반 필터링에 속하는 아이템기반 필터링방식으로 구분된다(Sarwar et al., 2001). 사용자기반 협업 필터링에 사용되는 기법으로는 피어슨(Pearson) 상관계수 또는 Constraint 피어슨 상관계수(Shardanand and Maes, 1995) 등이 있으며, 모델기반 협업 필터링 방법으로서 클러스터링(clustering), 베이지안 네트워크(Bayesian networks), 의사결정트리(decision tree), K-최대 근접이웃(K-nearest neighbor), 인공신경망(artificial neural networks) 등 여러가지 방법들이 제시되고 있다(Breese et al., 1998).

본 연구에서는 협업 필터링 방법으로서 피어슨(Pearson) 상관계수( $w$ )를 사용한다. 즉, 상관계수가 높은 이웃 고객들의 제품 선호정보를 토

대로 추천제품을 제시하는 방식을 이용한다. 예를 들어, 고객  $a$ 와 고객  $i$ 간의 피어슨 상관계수를  $w(a, i)$ 라고 할 때, 피어슨 상관계수는 식 (2)로 정의된다.

$$w(a, i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}} \quad (2)$$

$w_{a,i}$ : 고객  $a$ 와 고객  $i$ 간의 피어슨 상관계수

$v_{a,j}$ : 고객  $a$ 의 제품  $j$ 에 대한 선호도

$v_{i,j}$ : 고객  $i$ 의 제품  $j$ 에 대한 선호도

$\bar{v}_a$ : 고객  $a$ 의 전체 제품에 대한 평균 선호도

$\bar{v}_i$ : 고객  $i$ 의 전체 제품에 대한 평균 선호도

## 4. 실증분석 및 결과

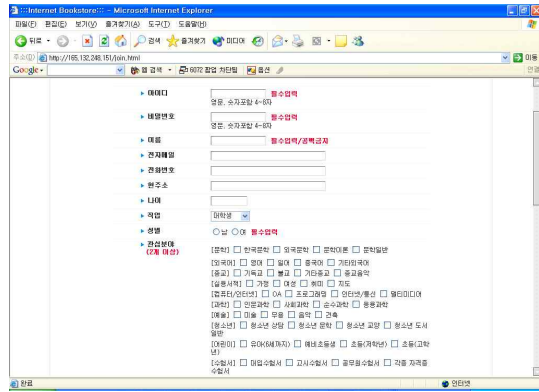
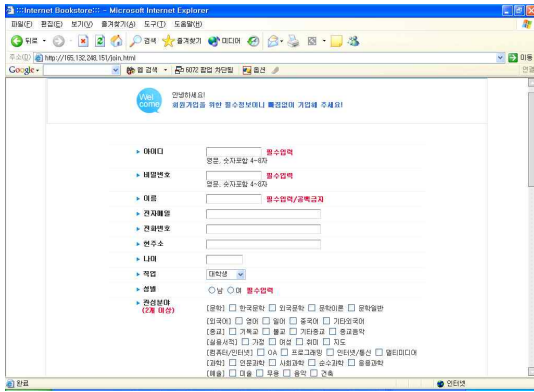
본 연구의 추천시스템의 타당성을 검증하기 위해 우선 가상 인터넷 서점을 구축한 후, 무작위로 추출된 표본 집단을 대상으로 실증분석 작업을 수행하였다. 이를 위한 실험대상 집단은 4년제 원주 및 부산지역 대학교 학부생들을 대상으로 조사되었으며, 분석대상 표본은 총 65명의 표본 수집자료가 실증분석에 사용되었다.

표본수집을 위해서는, 가상 인터넷 서점을 이용하기에 앞서서 우선 피실험자가 해당 서점에 접속하여 회원등록을 한 후, 선호도 설문조사에 참여하도록 하였으며, 그 다음에는 가상 서점에서 도서정보를 열람할 수 있도록 하였다. 가상 인터넷 서점의 웹사이트 화면은 [그림 2]와 같다.

<표 2>는 수집된 표본자료를 토대로 선호도 추정을 위한 회귀분석에 앞서서 선호도 영향 요인들간의 상관관계 및 선호도 영향요인과 선호도 평가 점수간의 상관관계를 분석한 결과를 보여



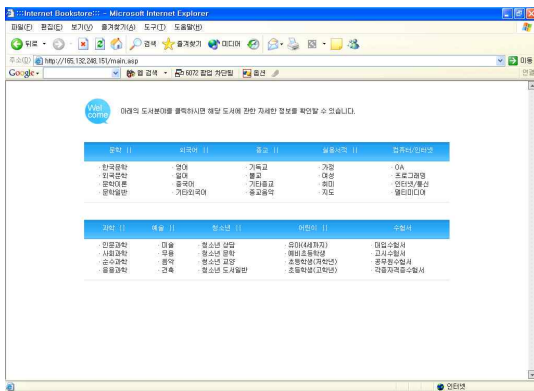
선호도 추정모형과 협업 필터링기법을 이용한 고객추천시스템



(a) 등록화면



(b) 선호도 설문조사 화면



(c) 홈페이지 화면

(d) 분야별 웹 페이지 화면 (예)

[그림 2] 가상 인터넷 서점의 웹페이지 화면

<표 2> 선호도 영향요인간의  
피어슨상관계수 분석결과

	장바구니 이용횟수	구매 횟수	관심서적 클릭횟수	관심유무	체류시간	선호도
장바구니 이용횟수	1.0000	-	-	-	-	-
구매 횟수	0.9569 (0.000)*	1.0000	-	-	-	-
관심서적 클릭횟수	0.8479 (0.000)*	0.8262 (0.000)*	1.0000	-	-	-
관심유무	0.4611 (0.000)*	0.4539 (0.000)*	0.4226 (0.000)*	1.0000	-	-
사이트 체류시간	0.4957 (0.000)*	0.5120 (0.000)*	0.6627 (0.000)*	0.2262 (0.000)*	1.0000	-
선호도	0.2956 (0.000)*	0.3037 (0.000)*	0.2667 (0.000)*	0.3442 (0.000)*	0.1351 (0.000)*	1.0000

\* : p < 1% 유의수준

주고 있다. 이 표에서 보는 바와 같이, 선호도 영향요인들 간에는 상관관계가 높게 나타나고 있다. 예를 들어, 장바구니 이용횟수, 구매횟수, 관심서적 클릭횟수간에는 매우 상관관계가 높게 나타났고, 사이트 체류시간과 관심서적 클릭 횟수 또한 높은 상관관계를 보여주고 있다.

이에 따라, 다중 회귀분석을 실시할 경우, 변수간의 상관관계가 높게 나타남에 따라 다중공선성의 문제가 발생할 가능성이 높게 존재한다. 이를 해결하기 위해 선택적 변수선정방법을 사용하여 다중 회귀분석을 실시하였다. <표 3>은

<표 3> 1차 선호도 추정모형 분석결과

변수	변수명	$\beta$	표준화 $\beta$	$t$ 값	$p$ 값
X1	관심유무	1.165	0.258	12.426	0.000*
X2	사이트 체류시간	0.178	0.042	-1.688	0.091**
X3	관심서적 클릭횟수	0.0407	0.053	1.303	0.193
X4	장바구니 이용횟수	0.144	0.057	-0.827	0.408
X5	구매 횟수	0.643	0.217	3.409	0.001*
A1	상수항	2.543		1003.036	0.000*

\* : p < 1% 유의수준, \*\* : p < 10% 유의수준

1차 다중 회귀분석결과를 보여주고 있다. 이 결과에서는 설명변수들간의 다중공선성의 문제로 말미암아 관심유무와 구매횟수를 제외한 나머지 변수들은 모두 유의적이지 않거나 회귀계수의 부호가 예상과 다른 상반된 결과를 초래하였다. 이에 따라, 다중공선성을 제거시키기 위한 2차 다중 회귀분석을 실시하였고, 이에 대한 분석결과는 <표 4>에서 보는 바와 같다. 즉, 고객의 선호도에 영향을 주는 변수들은 관심유무와 구매 횟수인 것으로 나타났으며, 표준화  $\beta$ 를 통해 알 수 있듯이 이 두 변수들이 고객 선호도에 영향을 미치는 중요도에 있어서 관심유무가 구매 횟수보다 더 높은 것으로 나타났다.

본 연구에서는 <표 4>의 분석결과를 토대로 선호도 추정모형을 도출한 후, 이 모형을 이용하여 고객 선호도를 추정한 다음, 협업 필터링을 사용하여 고객추천 성과를 분석하였다. 또한, 본 연구모형의 성과는 선호도 계산방식으로서 사전가중치에 의한 선호도를 이용하는 기존 추천시스템방식의 추천성과와 비교하여 그 성과 차이를 분석하였다.

<표 4> 2차 선호도 추정모형 분석결과

변수	변수명	$\beta$	표준화 $\beta$	$t$ 값	$p$ 값
X1	관심유무	1.172	0.259	12.617	0.000*
X5	구매 횟수	0.546	0.184	8.964	0.000*
A1	상수항	2.544		103.166	0.000*

0.000\* : p < 1% 유의수준

기존 연구에서 제시한 사전 가중치 부여방식(김중우 등, 1999)은 선호도 계산에 포함된 변수에 대한 가중치를 미리 전문가 등에 의해 고정적으로 부여하는 방식이다. 본 연구에서 제시하는 선호도 추정모형을 통한 가중치부여방식과 이러한 기존 방식의 성과를 비교하기 위해, 각

각의 선호도 계산방식을 협업 필터링을 통한 추천시스템에 적용하여 가중치부여방식에 따른 추천시스템의 성과를 비교 분석하였다. 비교분석을 위해 사용된 사전 가중치 부여방식의 선호도 가중치들은 모든 선호도 영향요인들에 대해 균등하게 가중치를 부여하는 방식을 적용하였다. 따라서, 본 연구의 실험결과는 이들 사전 가중치를 어떻게 차등적으로 부여하느냐에 따라서, 두 방식에 따른 협업 필터링의 성과의 차이는 서로 상이하게 나타날 것이다.

<표 5>는 본 연구에서 제안한 선호도 추정모형을 이용하여 선호도를 계산하는 방식의 협업 필터링의 성과가 기존 연구에서 제시된 선호도 가중치를 사전에 부여하는 방식을 통해 선호도를 계산하는 방식의 협업 필터링의 성과보다 더 높음을 보여주고 있다. 또한, 두 모형간의 성과 차이에 대한 유의성 검정인 대응표본 t검정에서도 선호도 추정모형을 이용한 협업 필터링기법의 성과가 유의적으로 더 높음을 보여주고 있다 (<표 6> 참조).

<표 5> 협업 필터링의 성과 비교

모형 구분	RMSE
사전가중치로서 단순평균 선호도를 이용한 협업 필터링기법	1.2510
선호도 추정모형을 이용한 협업 필터링기법	1.0950

<표 6> 협업 필터링기법 간의 성과차이검증 (대응표본 t검정)

	사전가중치로서 단순평균 선호도를 이용한 협업 필터링기법
선호도 추정모형을 이용한 협업 필터링기법	1.870 (0.062)*

\* : p < 10% 유의수준

## 5. 결론

본 연구는 새로운 선호도 추정모형을 제시하고, 이 모형을 통해 추정된 선호도를 이용하여 협업 필터링의 성과 제고 가능성을 살펴보았다. 그리고, 본 연구에서 제안한 선호도 추정모형을 통해 그동안 중요시 하지 않았던 선호도 측정에 영향을 주는 고객 행동 정보들간의 영향력을 실증적으로 분석하였고, 이를 통해 고객추천의 성과를 제고시킬 수 있는 방안들을 살펴보았다. 이러한 선호도 추정모형의 기대효과를 요약하면, 다음과 같다.

첫째, 본 선호도 추정모형을 통해 고객 선호도를 직접 추정해야 하는 어려움을 해소하고, 고객 선호도를 자동적으로 계산함으로써, 고객의 상품추천을 위한 선호도 계산의 효과성을 높일 수 있다. 둘째, 본 연구에서 제안한 선호도 추정모형은 그 동안 중요시 되어 오지 않았던 선호도 측정에 영향을 주는 웹 정보들간의 영향력을 최적화 모형을 통해 효과적으로 측정해 줌으로써, 고객추천시스템의 성과를 제고시킬 수 있다.

마지막으로 본 연구의 한계점은 다음과 같다.

우선, 실험대상 표본수의 부족과 설명변수의 제약, 가상 실험의 제약 등으로 말미암아 본 연구의 결과를 충분히 일반화시키기가 부족하다는 점이다. 예를 들어, 본 연구에서 수집한 자료들이 표본집단의 범위와 표본의 크기가 제약됨으로 인해서 이를 실제 온라인 서점에 적용하는 데에는 무리가 따를 수 있다는 점이다. 그리고, 선호도 추정모형의 다양한 접근방법을 고려하지 못하였다는 점이다. 예를 들어, 본 연구에서 고객 선호도는 선형함수로서 가정하고 있으나, 이론적으로 이러한 고객 선호도는 비선형함수가 더 일반적인 형태일 수 있다. 향후 연구에서

는 이러한 한계점들을 보완한 연구가 제시될 수 있을 것이다.

## 참고문헌

- [1] 김종우, 배세진, 이홍주, “협업 필터링 기반 개인화 추천에서의 평가자료의 희소정도의 영향”, *경영정보학 연구*, 14권 2호(2004), 131-149.
- [2] 김종우, 이경미, 김영국, 유관중, “인터넷 상점에서의 실시간 개인화된 광고 제공 기법”, *경영정보학연구*, 9권 4호(1999), 107-122.
- [3] 김종우, 이경미, “인터넷 상점에서 개인화 광고를 위한 장바구니 분석기법의 활용”, *경영과학*, 17권 3호(2000), 19-30.
- [4] 김진, 윤정섭, 조근식, “다중 정보여과 방법을 이용한 동적 정보 우선순위 결정”, *한국지능정보시스템학회 춘계학술대회*(2000), 323-332.
- [5] 박지선, 김택현, 류영석, 양성봉, “추천 시스템을 위한 2 way 협동적 필터링 방법을 위한 예측 알고리즘”, *정보과학회논문지: 소프트웨어 및 응용*, 29권 9호(2002), 669-675.
- [6] 이치훈, 정준, 정대진, 김용환, 이필규, “협력 및 내용기반 정보 여과방식을 이용한 추천 시스템설계”, *제14회 산, 학, 연 멀티미디어 산업기술 학술대회 논문집*(1999), 245-248.
- [7] 장근녕, 박유진, “인터넷쇼핑몰을 위한 개인화된 정보추천기법”, *Telecommunication Review*, 12권 1호(2002), 61-70.
- [8] 정경용, 류중경, 강운구, 이정현, “내용기반 여과와 협력적 여과의 병합을 통한 추천 시스템에서 조화 평균 가중치”, *정보과학회논문지: 소프트웨어 및 응용*, 30권 3호(2003), 239-250.
- [9] Balabanovic, M. and Y. Shoham, “Fab: Combing Content Based and Collaborative Recommendation”, *Communications of the ACM*, Vol.40, No.3(1997), 66-72.
- [10] Balabanovic, M., “An Adaptive Web Page Recommendation Service”, *Proceedings of the First International Conference on Autonomous Agents 1997: Marina del Rey, CA, USA*(1997), 378-385.
- [11] Breese, J.S, D. Heckerman, and C. Kadie, “Empirical Analysis of Predictive. Algorithms for Collaborative Filtering”, *Proceedings of the Fourteenth. Conférence on Uncertainty in Artificial Intelligence*(1998), 43-52.
- [12] BroadVision, “A BroadVision one to one White Paper”, White paper, BroadVision(1996).
- [13] Changchien, S. W. and T. Lu, “Mining Association Rules Procedure to Support On line Recommendation by Customers and Products Fragmentation”, *Expert Systems with Applications*, Vol.20, No.4(2001), 325-335.
- [14] Cho, Y. H., J. K. Kim, and S. H. Kim, “A Personalized Recommender System Based on Web Usage Mining and Decision Tree Induction”, *Expert Systems with Applications*, Vol.23, No.3(2002), 329-342.
- [15] Cohen, W. W., “Learning Rules that Classify e-mail”, *Proceedings of AAI Spring Symposium on Machine. Learning in Information Access*(1996), 18-25.
- [16] Goldberg, D., D. Nichols, B.M. Oki, and D. Terry, “Using Collaborative Filtering to Weave an Information Tapestry”, *Communications of the ACM*, Vol.35, No.12 (1992), 61-70.
- [17] Good, N., J. B. Schafer, J. A. Konstan, A. Borchers, B. Sarwa, J. Herlocker, and J. Riedl, “Combining Collaborative Filtering with Personal Agents for Better Recommendations”, *Proceedings of the Sixteenth National Conference on Artificial Intelligence, Orlando, FL*(1999), 439-446.
- [18] Herlocker, J. L., J. Konstan, A. Borchers, and J. Riedl, “An Algorithmic Framework for Performing Collaborative Filtering”, *Proceedings of the 1999 Conference on*

- Research and Development in Information Retrieval*(1999).
- [19] Hill, W., L. Stead, M. Rosenstein, and G. Furnas, "Recommending and Evaluating Choices in a Virtual Community of Use", *Proceedings of ACM CHI'95*, Denver, CO. (1995), 194-201.
- [20] Kim, J. W., B. H. Lee, M. J. Shaw, H. Chang, and M. Nelson, "Application of Decision Tree Induction Techniques to Personalized Advertisements on Internet Storefronts", *International Journal of Electronic Commerce*, Vol.5, No.3(2001), 45-62.
- [21] Kuo, Y. F. and L. S. Chen, "Personalization Technology Application to Internet Content Provider", *Expert Systems with Applications*, Vol. 21(2001), 203-215.
- [22] Lee, J., M. Podlaseck, E. Schonberg, and R. Hoch, "Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising", *Data Mining and Knowledge Discovery*, Vol.5 No.1/2(2001), 59-84.
- [23] Lieberman, H., "Autonomous Interface Agents", *Proceedings of the ACM Conference on Computers and Human Interface*, CHI 97, Atlanta, Georgia(1997), 67-74.
- [24] Maes, P. "Agent that Reduce Work and Information Overload", *Communications of the ACM*, Vol.37, No.7(1994), 30-40.
- [25] Mooney, R. J. and L. Roy, "Content based Book Recommending Using Learning for Text Categorization", *Proceedings of the Fifth ACM Conference on Digital Libraries* (2000), 195-204.
- [26] Oard, D. W. and J. Kim, "Implicit Feedback for Recommender Systems", *Proceedings of AAAI Workshop on Recommender Systems*, Madison, Wisconsin, USA(1998), 80-82.
- [27] Pazzani, M. J., "A Framework for Collaborative, Content Based and Demographic Filtering", *Artificial Intelligence Review*, Vol.13, No.5-6 (1999), 393-408.
- [28] Pazzani, M. J., J. Muramatsu, and D. Billsus, "Syskill & Webert: Identifying Interesting Web Sites", *Proceedings of the 13th National Conference on Artificial Intelligence*(1996).
- [29] Salton, G. and M. J. McGill, *An Introduction to Modern Information Retrieval*, NY: McGraw-Hill, 1983, 69-78.
- [30] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, "Item based Collaborative Filtering Recommendation Algorithms", *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong(2001), 285-295.
- [31] Sarwar, B. M., J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborating Filtering System", *Proceedings of the ACM Conference on Computer Supported Cooperative Work(CSCW'98)*, New York, ACM Press (1998), 345-354.
- [32] Schafer, J. B., J. A. Konstan, and J. Riedl, "Recommender System in E commerce", *Proceedings of the First ACM Conference on Electronic Commerce*(1999), 158-166.
- [33] Schafer, J. B., J. A. Konstan, and J. Riedl, "Electronic Commerce Recommender Applications", *Journal of Data Mining and Knowledge Discovery*, Vol.5, No.1-2 (2001), 115-152.
- [34] Shardanand, U. and P. Maes, "Social Information Filtering: Algorithms for Automating "Word of Mouth"", *Proceedings of CHI-95 Conference on Human Factors in Computing Systems*(1995), 210-217.
- [35] Wang, F. and H. Shao, "Effective Personalized Recommendation based on Time framed Navigation Clustering and Association mining", *Expert Systems with Applications*, Vol.27, No.3(2004), 365-377.

Abstract

## Customer Recommendation Using Customer Preference Estimation Model and Collaborative Filtering

Taeksoo Shin\* · Kun-Nyeong Chang\* · Youjin Park\*\*

This study proposed a customer preference estimation model for production recommendation and a method to enhance the performance of product recommendation using the estimated customer preference information. That is, we suggested customer preference estimation model to estimate exactly customer's product preference with his behavior. This model shows the relationship of customer's behaviors with his preferences. The proposed estimation model is optimized by learning the relative weights of customer's behavior variables to have an effect on his preference and enables to estimate exactly his preference.

To validate our proposed models, we collected virtual book store data and then made a comparative analysis of our proposed models and a benchmark model in terms of performance results of collaborative filtering for product recommendation. The benchmark model means a prior preference weighting model. The results of our empirical analysis showed that our proposed model performed better results than the benchmark model.

**Key words** : Collaborative Filtering, Customer Preference Estimation Model, Recommendation Systems

---

\* Dept. of Management Information Systems, College of Government and Business, Yonsei University

\*\* Dept. of Business Administration, The Graduate School, Yonsei University