

## 개인화된 제품 추천을 위한 개인과 그룹 행동에 기반한 고객 프로파일 모델 연구

### A study on the individual and group behavior based customer profile model for personalized products recommendation

박유진\*\*, 장근녕\*, 정유진\*\*

\*\* 연세대학교 경영학과 박사과정 (parkyoujin@yonsei.ac.kr)

\* 연세대학교 정경대학 경영학부 (knchang@yonsei.ac.kr)

\*\* 연세대학교 경영학과 박사과정 (eajjin@dragon.yonsei.ac.kr)

#### Abstract

일대일 마케팅을 실현하고 정보 과다 문제의 해결책으로 등장한 추천시스템의 다양한 기법을 적용하기 위해서는 고객의 관심 분야에 대한 정보인 고객 프로파일의 정의가 선행되어야 할 것으로 판단된다. 본 연구에서는 고객에게 개인화된 정보를 추천하기 위해 고객 개인의 행동과 그 고객이 속한 그룹의 행동 정보에 기반한 고객 프로파일 모델인 IGBCPM(Individual Group Behavior Customer Profile Model)을 제시한다.

#### 1. 서론

인터넷 기술의 발달로 e-커머스 업체들은 보다 낮은 원가를 투입하여 고객에 대한 정보를 간편하게 얻을 수 있게 되었으며, 고객의 개별적인 웹 경험을 관리하고 고객들과의 지속적인 커뮤니케이션 유지를 통해 고객 정보를 보다 높은 품질의 서비스 또는 제품으로 변화시킬 수 있게 되었다(Weng et al. 2004). 그러나 인터넷을 통해 얻을 수 있는 정보가 많아짐에 따라 웹 사용자들은 적절한 시간(right time)에 적절한 정보(right information)를 발견하기가 점점 어려워지는 정보 과다 문제(information overload problem)에 직면하게 되었다. 이러한 정보 과다 문제는 e-커머스 운영자들에게 고객에 대한 보다 편리하고 직관적인 제품 및 서비스 정보 제공의 필요성을 요구하고 있다.

이에 대한 주요 해결책으로 등장한 추천시스템은 예측된 선호 점수(likeness score) 또는 추천 제품 리스트 생성을 통해 고객들이 구매할 것 같은 제품들을 쉽게 발견할 수 있도록 도움을 주는 것에 데이터 분석 기법을 적용한 시스템이다(Sarwar et al. 1998). 현재 추천시스템은 도서, 영화, 음악, 뉴스 기사를 포함하는 다양한 아이템들을 추천하기 위해 Amazon, Yahoo, HP Shopping Village, Wal-Mart, Half.com, Musician's Friend 등의 유수의 전자상거래

사이트들에서 적용되고 있다(김종우 등, 2004; Ansari et al. 2000).

현재 보편적인 개인화된 추천기법에는 내용기반 기법과 협업필터링 기법이 있다. 내용기반 기법은 추천을 위해 제품 또는 서비스의 내용(content)과 고객의 관심을 상징하는 프로파일(profile)을 이용한다. 협업필터링 기법은 고객 사이에서 발견된 상관관계를 이용하여 비슷한 취향을 가진 고객들 간의 비교를 통해 새로운 고객이 흥미를 가지거나 구매할 가능성이 높은 정보를 예측하고 추천한다. 이러한 두 가지 접근법 외에 최근에는 웹 사용자의 행동 패턴을 발견하기 위해 데이터 마이닝 기법을 웹 데이터에 적용하는 프로세스인 web usage mining이 제시되고 있다.

이상과 같은 기법을 적용하기 앞서 고객들이 어떠한 제품 또는 제품 영역에 관심이 있는가를 나타내는 고객 프로파일을 정의하는 과정이 필요하다. Lee et al.(2003)은 개인화 작업에서 가장 중요한 이슈는 사용자의 선호를 예측하기 위한 개별 고객에 대한 계산 모델의 구축이라고 했다. 잘 정의된 고객 프로파일 모델은 내용기반 기법에서는 제품과 고객 선호의 일치여부를 판단하는데 활용될 수 있고, 협업 필터링 분석에서는 비슷한 취향이나 선호도를 가진 고객의 그룹화 과정에 활용될 수 있을 것이다. 본 연구의 목적은 웹상에서 얻을 수 있는 고객 행동 정보를 이용하여 보다 정확한 고객 프로파일 모델을 구축하고 이를 이용하여 개인화된 추천을 제공하는 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 추천시스템의 기법과 고객 프로파일 기법에 관한 기존 연구를 살펴보고, 3장에서는 개인과 그룹의 행동 정보에 기반한 새로운 고객 프로파일 모델을 제시한다. 4장에서는 제시된 모델을 평가하기 위한 실험 및 실험 결과를 제시하고, 5장에서는 결론을 제시한다.

#### 2. 관련 연구

## 2.1 추천시스템 기법

### 2.1.1 내용기반 기법

내용기반(content-based) 기법은 아이템(제품 또는 서비스)들의 내용을 분석하고 키워드, 구절, 특징과 같은 아이템과 관련하여 사용자의 관심을 상징하는 프로파일을 생성한다. 그리고 아이템의 내용과 사용자의 프로파일을 비교하고 그 아이템들이 사용자에게 관심을 끌 것인가를 평가한다(Min et al., 2005). 이 접근법은 고객의 프로파일에 대한 상품의 유사성에 기반하여 고객에게 제품을 추천하므로 내용 정보가 풍부한 기사 또는 웹 페이지와 같은 텍스트 정보의 추천에 유익하다. 내용기반 기법에서는 데이터 아이템의 표현과 프로파일 생성을 위한 사용자 선호의 묘사가 추천의 효과를 위한 핵심요소로(Lee et al. 2003) 사용자의 프로파일은 설문지, 평가된 항목, 또는 사용자 향해 정보(navigation information)에 의해 형성되어질 수 있다. 내용기반 기법을 적용한 추천시스템으로는 NewsWeeder, Inforfinder, 그리고 News Dude 등이 있다.

그러나 내용기반 기법은 텍스트 문서와 같이 특별한 유형의 정보에 대해서만 적용이 가능하고, 고객의 이전 경험과 유사한 제품만을 추천받을 수 있으며, 평가된 아이템들이 희소할 경우는 추천이 잘 이루어지지 못한다는 단점을 가지고 있다(Lee et al, 2001; Kwak et al. 2001).

### 2.1.2 협업 필터링 분석

협업 필터링(collaborative filtering) 분석은 비슷한 취향이나 선호도를 가진 고객들 간의 비교를 통해 새로운 고객이 흥미를 가지거나 구매할 가능성이 높은 정보를 예측하고 추천한다. 이를 위해서 취향이나 선호도가 유사한 최근접 이웃(nearest neighbor)을 발견해야 하는데 협업필터링에서는 유클리드 거리(Euclidean distance), 코사인 유사도, 상관관계 등과 같은 척도를 사용하여 사용자 간의 유사성을 산출해낸다. Goldberg et al.(1992)에 의해 개발된 Tapestry에서 협업필터링 기법이 처음으로 사용된 이래, Lotus Notes, GroupLens, Firefly, Sitemeter, PHOAKS, Grassroots 등과 같은 다양한 추천시스템들에서 협업필터링이 사용되고 있다.

그러나 협업필터링은 자료의 희소성(sparsity)과 자료의 확장성(scalability)이라는 문제점을 가지고 있다(Good et al. 1999). 이러한 문제점으로 인해 협업필터링에서는 사용자들에 의해 정보가 충분히 평가되기 전까지는 새로운 정보가 추천될 수 없으며, 사용자들이 늘어날수록 처리해야 하는 자료의 양이 기하급수적으로 증가하여 처리에 많은 시간을 요하게 된다. 협업필터링의 가장 큰 문제점인 희소성과 확장성을 해결하기 위해 분류(taxonomy)나 클러스터(cluster) 기법 등을 이용하여 분석에 이용되는 상품이나 사용자의 수를 줄이고, 묵시적(implicit)인 방법들을 이용하여 고객에게 제품에 대한 평가의 요구로 인한 번거로움과 이로 인한 평가 정보의 부족을 줄이고자 하는 연구가 수행되고 있다.

### 2.1.3 Web usage mining

웹 사용자가 특정한 웹 사이트를 탐색할 때 수집된 행동 정보는 사용자와 그 웹 사이트간의 상호작용을 의미한다. web usage mining은 데이터 마이닝 기법을 이용해서 웹 사용자의 행동 데이터(usage

data)를 분석하는 기법으로 웹 사용자에 대한 흥미 있는 이용 패턴을 발견하고 웹 사이트의 개선 및 고객에 대한 차별적인 서비스를 제공하는데 그 목적이 있다. 이것은 웹 개인화(web personalization)를 위한 주요 도구가 될 수 있기 때문에 e-커머스의 발전과 함께 그 중요성이 점점 커지고 있다.

현재, web usage mining에서는 클러스터링(clustering), 분류(classification), 연관규칙 발견(discovery of associations), 순차적 패턴 발견(sequential pattern discovery)과 같이 패턴발견에 관한 연구들이 활발하게 이루어지고 있다. 클러스터링은 데이터 집합을 그룹들 간에는 매우 다르고, 그룹 내 멤버 간에는 매우 유사한 그룹들로 나누는 것으로 Leader 알고리즘, BIRCH 알고리즘, Self-Organizing Maps 등이 있다. 분류는 클러스터링과 반대로 미리 정의된 클래스의 성격을 각 클래스의 속성(instance) 집합에 기반해서 구별하는 것으로 C4.5, CART, PIPPER 알고리즘 등이 있다. 연관규칙은 미리 정한 지지도(support)와 신뢰도(confidence)를 바탕으로 항목간 연관관계 규칙을 탐색하는 것으로 ARHP와 베이지안 네트워크(bayesian network) 등이 있다. 순차적 패턴이란 일정한 시간 동안 순차적으로 발생하는 거래를 말하는 것으로 WUM의 MINT 프로세서 모듈, SPSS의 클레멘타인 툴(Clementine tool) 등이 있다.

## 2.2 고객 프로파일링 기법

고객 프로파일(user profile)은 사용자의 욕구(needs)에 대한 구조화된 표현이다. 고객 프로파일링을 위한 기법에는 설문지와 인터뷰 등을 통해 고객의 지식을 습득한 후 고객들에 대한 공학적 통계 모델을 구축하고 가장 밀접한 모델에 고객을 일치시키는 방법인 지식기반(knowledge-based) 기법과 고객의 행동에서 유용한 패턴을 발견하기 위해 기계학습법을 사용하는 행위기반(behavior-based) 기법이 있다. 대부분의 추천시스템에서 고객 프로파일링은 행위기반 기법이 사용된다(Middleton et al. 2004).

지식기반 기법은 고객으로부터 직접 데이터를 입력받기 때문에 명시적 프로파일(explicit profile)이라고 할 수 있으며, 행위기반 기법은 고객의 피드백이나 행위 추적에 의해 프로파일링 되기 때문에 묵시적 프로파일(implicit profile)이라고 할 수 있다. 명시적 프로파일에서 고객들은 원하는 제품 또는 서비스를 얻기 위해 그들의 관심을 단어를 사용하여 표현해야 하는데, 이때 사용하는 단어가 너무 적거나 많은 경우라든지, 개념적인 내용(conceptual content)일 경우에는 관련 있는 문서와 일치되기 어렵다는 문제점이 있다(Fan et al. 2005).

전통적인 협업필터링에서는 웹 이용자들이 자신이 관심을 가지는 제품에 대하여 직접 점수를 입력(rating)하도록 하는 명시적인 방법을 이용하여 프로파일을 작성한다. 협업필터링을 이용한 대표적인 추천시스템인 GroupLens는 고객들이 숫자로 직접 입력한 영화에 대한 선호도를 고객 프로파일에 포함시키고 있고, Lee(2004)는 제품에 대한 관심 정도를 파악하기 위해 고객으로 하여금 직접 1-10까지의 값 중 하나로 제품에 대한 평가를 하도록 하였다.

묵시적인 방법을 이용한 고객 프로파일링 형성에는 구매 정보가 주로 사용되고 있다. Cho et al.(2005)은 어떠한 특정 기간에 대한 고객의 구매

처리 기록을 그 고객의 좋고 싫음을 묘사하는 고객 프로파일 형성을 위해 사용하였으며, 만약 고객이 그 제품을 구매하였다면 1, 그렇지 않다면 0의 값을 사용하여 프로파일을 표현하였다. Weng et al.(2004) 역시 고객의 구매 정보를 이진(binary) 정보로 표현하고 이들의 합과 고객이 구매한 제품 수와의 관계로 고객 프로파일을 작성하였다.

이외에 최근의 협업 필터링 기법에서는 고객의 평가를 직접 입력하는 방법과 더불어 고객의 웹 접속시간, 클릭 횟수, 웹 화면의 저장이나 출력 여부와 행동양식을 점수화하는 방법도 사용되고 있다.

### 3. 개인과 그룹 행동에 기반한 고객 프로파일 모델

앞에서 살펴본 Cho et al.(2005)과 Weng et al.(2004)의 연구에서는 이진 정보로 표현된 고객 개인의 구매 정보만을 이용하여 고객 프로파일을 작성하고 있다.

그러나 구매 여부 외에도 직·간접적으로 고객의 관심을 표명하는 행동들은 많이 존재한다. 예를 들어 웹사이트를 방문한 고객의 클릭, 북마크, 장바구니 담기, 인쇄하기 등은 간접적으로 고객의 관심을 표명하는 행동으로 볼 수 있다. 반면, 회원 가입시 고객이 입력하는 해당 제품에 대한 관심 여부는 직접적으로 고객의 관심을 표명하는 행동으로 볼 수 있다. 뿐만 아니라 개인의 정보만을 이용하여 고객 프로파일을 작성할 경우 현재 관심을 가지고 있는 정보만이 계속 추천되어 향후 흥미를 가지거나 구매할 가능성이 높은 정보들은 추천되지 못할 수도 있다. 그러므로 고객 개인의 구매 정보 외에도 고객과 유사한 취향을 가진 고객 그룹의 다양한 관심 표명 행위들이 고객의 프로파일 작성을 위해 반영되어져 할 것으로 판단된다.

한편, 고객 프로파일 작성을 위해 구매 정보를 0과 1이라는 이진 체계로 표현하는 것은 단순히 그러한 행동이 일어났다는 것에 대한 정보만을 나타낼 뿐 구체적인 강도에 대한 정보는 포함하고 있지 못하다. 그러므로 고객 프로파일 작성을 위해서는 다양한 관심표명 행위들의 실제적인 행동 정도가 반영되어야 할 것으로 판단된다.

여기에서는 웹 상에서 발생할 수 있는 다양한 관심표명 행위 중에서 고객 선호와 밀접한 관련이 있는 정보로 판단되는 고객 개인과 고객 그룹의 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심분야 체크여부에 대한 실제 정도에 대한 정보를 이용한 고객 프로파일 모델을 제시한다.

#### 3.1 제품 프로파일 모델

제품 프로파일(product profile) 모델은 제품의 특성을 정리해 놓은 것으로, 개인화된 추천을 위한 고객 프로파일과의 유클리드 거리 계산에 이용된다. 제품 m에 대한 제품 프로파일은 다음과 같이 정의된다.

$$PP_m = (f_m^{ij}, i = 1, \dots, I, j = 1, \dots, K_j) \quad m = 1, \dots, M$$

M: 제품의 총 수  
 I: 제품 특징의 수  
 $K_j$ : 제품 특징 j를 구성하는 특징값의 수 ( $i = 1, \dots, I$ )  
 $f_m^{ij}$ : 제품 m이 제품 특징 i, j(제품 특징 i, 특징값 j)를 가지고 있으면 1, 그렇지 않으면 0 ( $i = 1, \dots, I, j = 1, \dots, K_j$ )

예를 들어 가방의 제품 특징을 색상, 재질, 용도 등의 3가지로 구분하고 (즉,  $f_3$ ,  $f_1$ (색상),  $f_2$ (재질),  $f_3$ (용도)), 색상이라는 특징의 구체적인 특징값으로는 검정, 노랑, 파랑(즉,  $K_1=3$ ,  $f_1$ (검정),  $f_2$ (노랑),  $f_3$ (파랑)), 재질이라는 특징의 구체적인 특징값으로는 가죽과 천(즉,  $K_2=2$ ,  $f_1$ (가죽),  $f_2$ (천)), 용도라는 특징의 구체적인 특징값으로는 서류용과 여행용(즉,  $K_3=2$ ,  $f_1$ (서류용),  $f_2$ (여행용))으로 정의한다고 하자. 이 때 가죽으로 만든 노란색 서류 가방의 제품 프로파일은 다음과 같이 정의된다.

$$PP_m = (f_m^{11}, f_m^{12}, f_m^{13}, f_m^{21}, f_m^{22}, f_m^{31}, f_m^{32}) = (0, 1, 0, 1, 0, 1, 0)$$

### 3.2 개인과 그룹 행동 기반 고객 프로파일

#### 3.2.1 개인 관심수준 계산

개인 프로파일 모델은 고객 개인의 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심분야 체크여부를 이용하여 각 제품 특징에 대한 개인의 관심 수준을 분석한 것으로, 우선 특정 고객 A의 관심 수준을 계산해 내고, 이를 전체 고객에 대한 상대적인 수준으로 계산하여 만들어진다.

고객 A의 제품 특징 i, j(제품 특징 i, 특징값 j)에 대한 가중 관심 수준을 나타내는 WII(Weighted Individual Interest)는 다음과 같이 구해진다.

$$WII_A^{ij} = \frac{b_A^{ij}}{B_A} + \frac{c_A^{ij}}{C_A} + \frac{bs_A^{ij}}{BS_A} + \frac{p_A^{ij}}{P_A}$$

$WII_A^{ij}$ : 고객 A의 제품 특징 i, j에 대한 가중 관심

$B_A$ : 고객 A가 구매한 제품의 총 수

$b_A^{ij}$ : 제품 특징 i, j를 만족하는 제품에 대한 고객 A의 구매 횟수

$C_A$ : 고객 A가 클릭한 제품의 총 수

$c_A^{ij}$ : 제품 특징 i, j를 만족하는 제품에 대한 고객 A의 클릭 횟수

$BS_A$ : 고객 A가 장바구니 담기한 총 수

$bs_A^{ij}$ : 제품 특징 i, j를 만족하는 제품에 대한 고객 A의 장바구니 담기 횟수

$P_A$ : 고객 A가 관심분야로 체크한 제품의 총 수

$p_A^{ij}$ : 제품 특징 i, j를 만족하는 제품에 대한 고객 A의 관심분야 체크여부

WIRI(Weighted Individual Relative Interest)는 제품 특징 i, j에 대한 고객 A의 관심 수준을 전체 고객과 비교하여 상대적으로 계산한 값으로 다음과 같이 구해진다.

$$WIRI_A^{ij} = \frac{WII_A^{ij}}{\frac{1}{|S|} \sum_{s \in S} WII_s^{ij}}$$

S: 모든 고객의 집합

### 3.2.2 그룹 관심수준 계산

고객A와 유사한 행동 특성을 보이는 고객들로 이루어진 그룹A를 발견하기 위해 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심분야 체크여부와 같은 고객들의 관심표명 행동을 분석하였다. 즉, 고객A가 관심표명을 한 제품들에 대해 높은 관심 표명 행동을 한 고객들이 그룹A로 분류된다.

고객 A와 유사한 행동 특성을 보이는 고객들의 그룹인 그룹A가 형성되고 나면, 그룹 전체의 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심분야 체크여부를 이용하여 각 제품 특징에 대한 그룹의 관심 수준을 분석한다. 그룹 관심수준의 계산 과정은 개인 관심수준 계산 과정과 동일하게 이루어진다. 우선 고객A가 속한 그룹A 전체의 관심 수준을 계산해 내고, 이를 전체 그룹에 대한 상대적인 수준으로 계산한다.

그룹A의 제품 특징  $i, j$ (제품 특징  $i$ , 특징값  $j$ )에 대한 가중 관심 수준을 나타내는 WGI(Weighted Group Interest)는 다음과 같이 구해진다.

$$WGI_A^{ij} = \frac{b_{Ag}^{ij}}{B_{Ag}} + \frac{c_{Ag}^{ij}}{C_{Ag}} + \frac{bs_{Ag}^{ij}}{BS_{Ag}} + \frac{p_{Ag}^{ij}}{P_{Ag}}$$

$WGI_A^{ij}$ : 그룹 A의 제품 특징  $i, j$ 에 대한 가중 관심

$B_{Ag}$ : 그룹 A가 구매한 제품의 총수

$b_{Ag}^{ij}$ : 제품 특징  $i, j$ 를 만족하는 제품에 대한 그룹 A의 구매 횟수

$C_{Ag}$ : 그룹 A가 클릭한 제품의 총수

$c_{Ag}^{ij}$ : 제품 특징  $i, j$ 를 만족하는 제품에 대한 그룹 A의 클릭 횟수

$BS_{Ag}$ : 그룹 A가 장바구니 담기한 총수

$bs_{Ag}^{ij}$ : 제품 특징  $i, j$ 를 만족하는 제품에 대한 그룹 A의 장바구니 담기 횟수

$P_{Ag}$ : 그룹 A가 관심분야로 체크한 제품의 총수

$p_{Ag}^{ij}$ : 제품 특징  $i, j$ 를 만족하는 제품에 대한 그룹 A의 관심분야 체크여부

WGR(Weighted Group Relative Interest)는 제품 특징  $i, j$ 에 대한 그룹A의 관심 수준을 전체 그룹과 비교하여 상대적으로 계산한 값으로 다음과 같이 구해진다.

$$WGR_A^{ij} = \frac{WGI_A^{ij}}{\frac{1}{|P|} \sum_{p \in P} WGI_p^{ij}}$$

P: 모든 그룹의 집합

### 3.2.3 고객 프로파일 모델

고객A의 행동정보와 그룹A의 행동정보가 반영된 개인과 그룹 행동기반 고객 프로파일 모델은 다음

과 같이 계산된 WIGRI(Weighted Individual Group Relative Interest)를 통해 표현된다.

$$WIGRI_A^{ij} = (\alpha_1 \times WIRI_A^{ij}) + (\alpha_2 \times WGR_A^{ij})$$

$\alpha_1$ : WIRI에 대한 가중치

$\alpha_2$ : WGR에 대한 가중치

$$(\alpha_1 + \alpha_2 = 1)$$

$\alpha_1$ 과  $\alpha_2$ 는 WIGRI에 대한 WIRI와 WGR의 상대적인 가중치로, 본 연구에서는  $\alpha_1, \alpha_2$ 를 각각 0.5로 설정하였다. 이것은 WIGRI를 계산하기 위해 WIRI와 WGR를 동일한 비율로 사용한다는 것을 의미한다.

고객A의 개인과 그룹 행동기반 고객 프로파일 IGBCP(Individual Group Behavior based Customer Profile)는 다음과 같이 표현되어진다.

$$IGBCP_A = (WIGRI_A^{ij}, i = 1, \dots, I, j = 1, \dots, K_i)$$

이상에서 보는 바와 같이 본 연구에서 제시하는 고객 프로파일 모델은 각 제품 특징  $i, j$ 에 대한 고객 개인과 고객 개인이 속한 그룹의 클릭 횟수와 구매 횟수, 장바구니 담기 횟수, 그리고 관심분야 체크여부를 반영한 고객의 상대적인 관심의 수준이다.

### 3.3 제품 추천 모델

여기에서는 앞에서 제시한 제품 프로파일 모델과 개인과 그룹 행동기반 고객 프로파일 모델을 이용한 개선된 제품추천 모델을 제시한다. 먼저 개인화된 정보를 추천하기 위해 제시된 고객 프로파일과 제품 프로파일간의 유사성을 계산한다. 고객A와 제품 m간의 유사성은 유클리드 거리 계산법을 이용하여 다음과 같이 계산되어진다.

$$R_{Am} = \sqrt{\sum_j (WIGRI_A^{ij} - f_m^{ij})^2}$$

전체 제품 m에 대한  $R_{Am}$ 을 계산하고,  $R_{Am}$ 이 큰 순서대로 제품을 추천한다.

## 4. 실험 및 결과 분석

### 4.1 실험 데이터

추천시스템의 성능 평가를 위해 실제 운영되고 있는 상업용 인터넷 쇼핑몰의 데이터와 MovieLens<sup>1)</sup>나 EachMovie<sup>2)</sup>와 같이 일부 연구소에서 제공하

- 1) GroupLens Research Project에서 7개월 동안 <http://www.grouplens.org>를 통해 수집한 데이터로, 943명의 고객이 1,682개의 영화에 대해 1-5점으로 총 100,000개를 평가하였다.
- 2) Compaq Systems Research Center에서 협업 필터링 알고리즘 실험을 위해 18개월 동안 <http://research.compaq.com/SRC/eachmovie>를 통해 수집한 데이터로 72,916명의 고객이 1,628개의 영화에 대해 0-1(0.2간격)점으로 총 2,811,980개를 평가하였다.

고 있는 데이터를 사용한다. 그러나 상업용 인터넷 쇼핑몰에서 발생하는 고객 데이터는 수집하기가 그리 간단하지 않으며, MovieLens나 EachMovie와 같은 데이터는 본 연구에서 사용하기에 적절하지 않은 것으로 판단된다. 따라서 Windows XP Professional, IIS 6.0 환경에서 ASP(Active Server Pages)를 개발언어로 사용하고 Microsoft Access를 데이터베이스로 사용하는 도서정보 추천시스템을 <표 1>과 같은 도서 분류 체계에 따라 구축하고 실험을 통해 데이터를 생성하였다.

제시한 기법의 효과를 분석하기 위해 Weng et al.(2004)이 제시한 고객 프로파일 모델(Weng Model)을 적용한 도서정보 추천시스템과 고객 개인의 행동 정보를 반영하는 개인 행동 기반 고객 프로파일 모델(BCPM: Individual Behavior-based Customer Profile Model)을 적용한 도서정보추천시스템, 그리고 고객 개인과 그 고객이 속한 그룹의 행동정보를 반영하는 개인과 그룹 행동 기반 고객 프로파일 모델(IGBCPM: Individual Group Behavior-based Customer Profile Model)을 적용한 도서정보 추천시스템을 구축하였다.

구축된 도서정보 추천시스템들의 피실험자로 인터넷 사용이 익숙한 74명의 대학생 및 일반인을 선정하였으며, 충분한 데이터 수집을 위해 10회 이상 도서를 구입하도록 하였다. 단, 도서 1회 구입시 도서 구입 의사결정을 위한 도서정보 열람은 4회 이상하도록 하였는데, 이 때 열람되는 도서 정보가 모두 동일해서는 안된다. 이상과 같은 실험 과정을 통해 최종적으로 58명의 피실험자에 대한 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 그리고 관심분야 정보가 수집되었고 이 데이터를 분석 데이터로 이용하였다.

<표 1> 도서 분류 체계

구분	항목
문학	한국문학, 외국문학, 문학이론, 문학일반
외국어	영어, 일어, 중국어, 기타 외국어
종교	기독교, 불교, 기타종교, 종교음악
실용서적	가정, 여성, 취미, 지도
컴퓨터/인터넷	OA, 프로그래밍, 인터넷/통신, 멀티미디어
과학	인문과학, 사회과학, 순수과학, 응용과학
예술	미술, 무용, 음악, 건축
청소년	청소년 상담, 청소년 문학, 청소년 교양, 청소년 도서일반
어린이	유아(6세까지), 예비초등학생, 초등학생(저학년), 초등학생(고학년)
수험서	대입수험서, 고시수험서, 공무원수험서, 각종자격증수험서

#### 4.2 평가 지표

추천시스템의 성능을 평가하기 위해 정보검색 시스템 평가 지표로 가장 잘 알려진 방법이며, Billsus et al.(1998), Basu et al.(1008), Sarwar et al.(1998) 등에 의해 추천시스템의 평가 지표로도 사용되어 오고 있는 정확률(precision)과 재현률(recall), 그리고 F1을 사용하였다. F1은 정확률과 재현률에 동등한

가중치를 부여하여 하나의 평가항목으로 결합한 것이다.

$$P = \frac{N_{BR}}{N_R}$$

$$R = \frac{N_{BR}}{N_B}$$

$$F1 = \frac{2PR}{P+R}$$

P: 정확률

R: 재현률

$N_{BR}$ : 추천된 도서 중 실제 구매된 도서의 수

$N_R$ : 추천된 도서의 수

$N_B$ : 구매된 도서의 수

#### 4.3 결과 분석

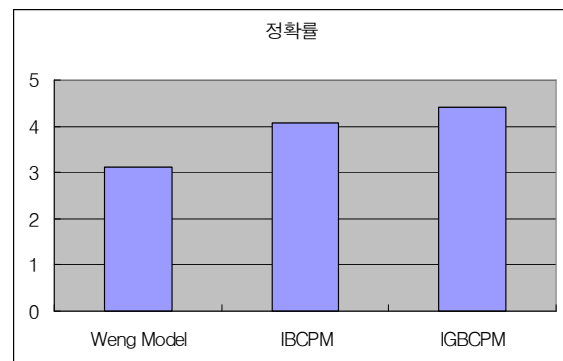
Weng 모델과 IBCPM, 그리고 IGBCPM에서 추천되는 도서 정보의 크기가 5일 때의 시스템별 평균 정확률과 재현률, 그리고 F1의 비교 결과가 <표 2>와 [그림 1]~[그림 3]에 정리되어 있다.

추천된 도서의 수와 추천된 도서 중 실제 구매된 도서의 수에 대한 비율을 나타내는 정확률을 비교해 놓은 <표 2>와 [그림 1]에서 보는 바와 같이 각 시스템의 정확률은 IGBCPM>IBCPM>Weng Model 순서로 나타났다. 이 결과를 통해 추천시스템의 정확률을 향상시키기 위한 고객 프로파일은 단순히 2진 형태로 표현된 구매 정보만을 이용하여 형성된 것 보다는 고객 개인의 다양한 행동 정보와 더불어 유사한 고객 그룹의 다양한 행동 정보의 실제 정도를 이용하여 형성된 것이라는 것을 알 수 있다.

<표 2> 시스템 별 평균 정확률, 재현률, F1

	정확률	재현률	F1
Weng Model	3.109	8.2	4.294
IBCPM	4.087	16.8	6.5
IGBCPM	4.408	21.242	7.291

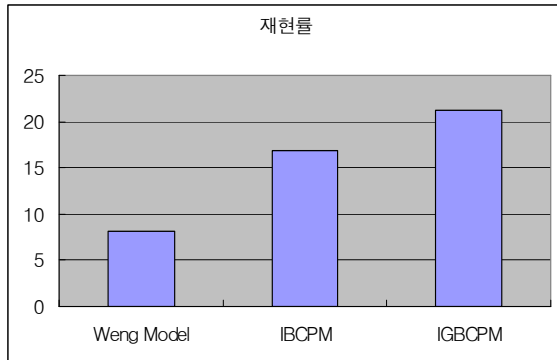
[그림 1] Weng, IBCPM, IGBCPM의 정확률 비교



고객이 구매한 도서의 수와 추천된 도서 중 실제

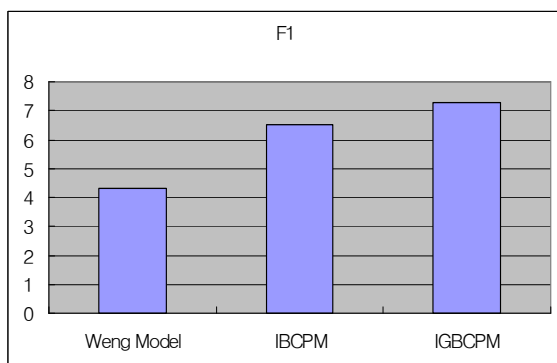
구매된 도서의 수에 대한 비율을 나타내는 재현률을 비교해 놓은 <표 2>와 [그림 2]에서 보는 바와 같이 각 시스템의 재현률은 IGBCPM>IBCPM>Weng Model 순서로 나타났다. 이 결과를 통해 추천시스템의 재현률을 향상시키기 위한 고객 프로파일은 단순히 2진 형태로 표현된 구매 정보만을 이용하여 형성된 것 보다는 고객 개인의 다양한 행동 정보와 더불어 유사한 고객 그룹의 다양한 행동 정보의 실제 정도를 이용하여 형성된 것이라는 것을 알 수 있다.

[그림 2] Weng, IBCPM, IGBCPM의 재현률 비교



정확률과 재현률을 하나의 평가항목으로 결합한 F1을 비교해 놓은 <표 2>와 [그림 3]에서 보는 바와 같이 각 시스템의 F1은 IGBCPM>IBCPM>Weng Model 순서로 나타났다. 이 결과를 통해 단순히 2진 형태로 표현된 구매 정보만을 이용하여 형성된 고객 프로파일을 이용한 도서정보의 추천보다는 고객 개인의 다양한 행동 정보와 더불어 유사한 고객 그룹의 다양한 행동 정보의 실제 정도를 이용하여 형성된 고객 프로파일을 이용한 추천시스템이 개인화된 정보추천에 보다 유용하다는 것을 알 수 있다.

[그림 3] Weng, IBCPM, IGBCPM의 F1 비교



## 5. 결론

고객에게 개인화된 정보를 제공하기 위해서는 우선 고객이 어떠한 정보를 선호하는가를 나타내는 고객 프로파일을 정의하는 과정이 필요하다. 본 연구에서는 웹 상에서 발생할 수 있는 다양한 관심표명 행위 중에서 고객 선호와 밀접한 관련이 있는

정보로 판단되는 고객 개인과 고객 그룹의 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심분야 체크 여부에 대한 실제 정도에 대한 정보를 이용한 고객 프로파일 모델을 제시하였다.

이 기법은 제품의 특징 정보와 고객 개인과 그 고객과 유사한 행동 정보를 가지는 그룹의 행동 정보에 기반하고 있다. 이전의 연구들이 고객 프로파일을 작성하기 위해 2진 형태로 표현된 개인의 구매 정보만을 사용하고 있는데 반해 본 기법은 고객 개인과 그룹의 실제 구매 횟수, 클릭 횟수, 장바구니 담기 횟수와 고객이 직접 입력한 관심분야에 대한 정보를 사용하였다. 이러한 정보를 토대로 제품 특징들에 대한 상대적인 가중 관심 수준을 도출해 낸 후 이것을 제품 특징별로 정리한 개인과 그룹 행동기반 고객 프로파일 모델 IGBCPM(Individual Group Behavior-based Customer Profile Model)을 정의하였다. 마지막으로 제품의 추천은 유클리드 거리 계산법을 사용하여 고객 프로파일과 미리 정의된 제품 프로파일간의 유사성을 계산하여 유사성이 높은 제품을 추천해 주었다.

제시한 기법의 효과를 분석하기 위해 Weng et al.(2004)이 제시한 고객 프로파일 모델(Weng Model)을 적용한 도서정보 추천시스템과 고객 개인의 행동 정보를 반영하는 개인 행동 기반 고객 프로파일 모델(IBCPM: Individual Behavior-based Customer Profile Model)을 적용한 도서정보 추천시스템, 그리고 고객 개인과 그 고객이 속한 그룹의 행동정보를 반영하는 개인과 그룹 행동 기반 고객 프로파일 모델(IGBCPM: Individual Group Behavior-based Customer Profile Model)을 적용한 도서정보 추천시스템을 구축하였다. 각각의 시스템에 대하여 정확률, 재현률, 그리고 F1을 평가한 결과 세 가지 평가 지표 모두 IGBCPM>IBCPM>Weng Model 순으로 나타나 단순히 2진 형태로 표현된 구매 정보만을 이용하여 형성된 고객 프로파일을 이용한 도서정보의 추천보다는 고객 개인의 다양한 행동 정보와 더불어 유사한 고객 그룹의 행동 정보의 실제 정도를 이용하여 형성된 고객 프로파일이 개인화된 정보추천 시스템에 보다 유용하다는 것을 알 수 있었다.

본 연구는 고객의 선호를 나타내는 고객 프로파일을 정의할 때 단순히 2진 형태로 표현된 개인의 구매 정보만을 사용하지 않고 개인, 개인과 유사한 그룹의 구매 횟수, 클릭 횟수, 장바구니 담기 횟수와 같은 웹 상에서의 행동들과 고객이 직접 제공한 관심 분야에 대한 정보를 이용한 새로운 고객 프로파일을 제시하였다는데 그 의의가 있다.

본 논문에서 제시한 고객 프로파일 모델을 2장에서 살펴보았던 다양한 추천 기법에 활용한다면 고객이 원하거나 또는 관심을 가질 가능성이 높은 제품을 보다 정확하게 추천해 줄 수 있을 것이다. 이것은 제품 정보 검색에 소비되는 고객의 시간과 노력을 줄여줄 것이고, 나아가 해당 사이트에 대한 고객의 만족도와 충성도를 향상시키게 될 것이다. 뿐만 아니라 고객이 관심을 가질만한 적절한 제품의 추천을 통해 구매 욕구를 자극시켜 새로운 수요 창출을 유발할 수 있을 것이다.

그러나 본 연구에서는 해당 고객과 다른 고객에 대한 행동정보의 일치 여부만을 고려하여 유사한 고객 그룹을 생성하였기 때문에 보다 합리적인 고객 그룹화가 이루어지지 못하였다는 데 그 한계가 있다. 그러므로 향후에는 클러스터링 기법이나 피

어슨 상관계수, 코사인 유사도와 같은 기법을 적용한 고객 그룹화 작업이 보강되어야 할 것으로 판단된다. 그리고 클릭 횟수, 구매 횟수, 장바구니 담기 횟수 등과 같은 다양한 관심표명 행동이 고객의 관심 수준에 어떠한 영향을 미치는 지를 나타내는 가중치에 대한 분석을 통해 고객의 행동정보에 따라 변화될 수 있는 최적 가중치의 선정에 대한 연구도 이루어져야 할 것으로 판단된다.

## 참고 문헌

- 김종우, 배세진, 이홍주 (2004), 협업 필터링 기반 개인화 추천에서의 평가자료의 희소 정도의 영향, *경영정보학연구*, 14(2), 131-149.
- Ansari, A., Essegai, S., and Kohli, R. (2000), Internet Recommendation Systems, *Journal of Marketing Research*, 37(3), 363-375.
- Sarwar, B.M., Konstan, J.A., Borchers, A., Herlocker, J., Miller, B., and Riedl, J. (1998), Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System, *Proceedings of CSCW 98*, Seattle, Washington, 345-354.
- Basu, C., Hirsh, H., and Cohen, W. (1998), Recommendation as classification: using social and content-based information in recommendation, *Proceedings of the 1998 workshop on recommender systems*, Menlo Park, CA: AAAI Press, 714-720.
- Lee, C.H., Kim, Y.H., and Rhee, P.K. (2001), Web personalization expert with combining collaborative filtering and association rule mining technique, *Expert Systems with Applications*, 21(3), 131-137.
- Billsus, D., and Pazzani, M.J. (1998), Learning Collaborative Information Filters, *Proceedings of ICML*, 46-53.
- Goldberg, D., Nichols, D., Oki, B.M., and Terry, D. (1992), Using Collaborative Filtering to Weave an Information Tapestry, *Communications of the ACM*, 35(12), 61-70.
- Kwak, M., and Cho, D.S. (2001), Collaborative filtering with automatic rating for recommendation, *Proceedings of ISIE 2001 IEEE international symposium on industrial electronics*, 625-628.
- Good, N., Schafer, J.B., Konstan, J.A., Borchers, A., Sarwa, B., Herlocker, J., and Riedl, J. (1999), Combining collaborative filtering with personal agents for better recommendations, *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, Orlando, FL, 439-446.
- Middleton, S.E., Shadbolt, N.R., and de Roure, D.C. (2004), Ontological User Profiling in Recommender Systems, *ACM Transactions on Information Systems*, 22(1), 54-88.
- Min, S.H., and Han, I. (2005), Detection of the customer time-variant pattern for improving recommender systems, *Expert Systems with Application*, 28(2), 189-199.
- Weng, S.S., and Liu, M.J. (2004), Feature-based recommendations for one-to-one marketing, *Expert Systems with Applications*, 26(4), 493-508.
- Fan, W., Gordon, M.D., and Pathak, P. (2005), Effective profiling of consumer information retrieval needs: a unified framework and empirical comparison, *Decision Support Systems*, 40(2), 213-233.
- Lee, W.P. (2004), Applying domain knowledge and social information to product analysis and recommendations: an agent-based decision support system, *Expert Systems*, 21(3), 138-148.
- Lee, W.P., and Yang, T.H. (2003), Personalizing information appliances: a multi-agent framework for TV programme recommendations, *Expert Systems with Applications*, 25(3), 331-341.
- Cho, Y.B., Cho, Y.H., and Kim, S.H. (2005), Mining changes in customer buying behavior for collaborative recommendations, *Expert Systems with Applications*, 28(2), 359-369.