

개인화된 제품 추천을 위한 고객 행동 기반 고객 프로파일링 기법

박유진* · 정유진* · †장근녕*

Customer Behavior Based Customer Profiling Technique for Personalized Products Recommendation

Youjin Park* · Eaujin Jung* · Kun-Nyeong Chang*

■ Abstract ■

In this paper, we propose a customer profiling technique based on customer behavior for personalized products recommendation in Internet shopping mall. The proposed technique defines customer profile model based on customer behavior information such as click data, buying data, market basket data, and interest categories. We also implement CBCPT(customer behavior based customer profiling technique) and perform extensive experiments. The experimental results show that CBCPT has higher MAE, precision, recall, and F1 than the existing other customer profiling technique.

Keyword : Customer Profiling Technique, Personalized Products Recommendation, Recommender System

1. 서 론

인터넷 기술의 발달로 e-커머스 업체들은 보다 낮은 원가를 투입하여 고객에 대한 정보를 간편

하게 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

논문접수일 : 2006년 08월 10일 논문게재확정일 : 2006년 10월 30일

* 연세대학교 원주캠퍼스 경영학과

† 교신저자

[34]. 그러나 인터넷을 통해 얻을 수 있는 정보가 많아짐에 따라 웹 사용자들은 적절한 시간(right time)에 적절한 정보(right information)를 발견하기가 점점 어려워지는 정보과다 문제(information overload problem)에 직면하게 되었다.

이에 대한 주요 해결책으로 등장한 추천시스템은 예측된 선호 점수(likeness score) 또는 추천 제품 리스트 생성을 통해 고객들이 구매할 제품들을 쉽게 발견할 수 있도록 도움을 주는 시스템이다 [30]. 추천시스템에서 사용되어지고 있는 개인화된 추천 기법은 내용기반 필터링과 협업 필터링으로 나누어지는데[24, 26, 27], 내용기반 필터링은 고객의 관심 분야에 대한 제품의 유사성에 기반하여 제품을 추천하는 반면, 협업 필터링은 이웃(neighbor)으로 알려진 고객 그룹의 관심 분야에 기반하여 제품을 추천한다. 현재 추천시스템은 도서, 영화, 음악, 뉴스 기사를 포함하는 다양한 아이템들을 추천하기 위해 Amazon, Yahoo, HP Shopping Village, Wal-Mart, Half.com, Musician's Friend 등과 같은 유수의 전자상거래 사이트들에 적용되고 있다 [1, 3].

개인화된 제품을 추천하기 위해서는 고객이 어떠한 제품 또는 제품 영역에 관심이 있는가를 나타내는 고객 프로파일(customer profile)의 정의가 선행되어야 한다. Lee 등[22]과 Fan 등[9]은 개인화 작업에서 가장 중요한 이슈는 사용자의 선호를 예측하기 위한 개별 고객의 계산 모델 구축이라고 하였고, Montaner 등[28]은 추천시스템의 성공은 정확한 프로파일의 구축에 달려있으며, 정확한 프로파일은 내용기반 필터링에서는 추천의 적절성(appropriate)을 보장해주고, 협업 필터링에서는 유사한 프로파일을 가진 고객들이 정말로 유사한지를 보장해 준다고 하였다. 그러므로 추천 기법들을 적용하기에 앞서 고객 프로파일을 정확하게 정의하는 것은 중요한 작업이라 할 수 있다.

고객 프로파일링을 위해 필요한 고객의 선호 정보를 수집하는 방법은 크게 관심 정보를 고

객으로부터 직접 입력 받을 것인가? 아니면 간접적으로 예측할 것인가?로 분류할 수 있으며, 최근의 추천시스템에서는 정보 수집의 편리함과 데이터 분석 기술의 발전으로 웹상에서의 고객 행동 분석을 통해 고객 프로파일이 정의되고 있다.

고객 행동분석을 통한 고객 프로파일 정의와 관련된 기존의 연구들에서는 웹상에서의 고객의 다양한 관심 표명 행동 가운데 주로 구매 정보를 이용하여 고객 프로파일을 정의하고 있다[8, 31, 32, 34]. 구매 정보가 고객의 관심을 나타내는 주요 지표로 사용될 수 있으나[7], 고객의 관심을 보다 정확하게 예측하기 위해서는 구매 정보 이외에 웹상에서의 다양한 행동들에 대한 정보들을 반영할 필요가 있다. 그러므로 본 논문에서는 고객의 선호를 보다 잘 반영하는 고객 프로파일을 정의하기 위해 웹상에서 발생할 수 있는 고객의 다양한 직·간접적 관심 표명 행동 중에서 고객의 선호와 가장 밀접한 관련이 있는 정보로 판단되는 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 그리고 회원가입시 고객으로부터 직접 입력되는 관심 분야 정보를 반영한 고객 프로파일링 기법을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 추천 기법과 고객 프로파일링에 관한 기존연구를 살펴보고, 3장에서는 고객 행동기반 고객 프로파일링 기법과 제품 추천 기법을 제시한다. 4장에서는 제시된 기법을 평가하기 위한 실험결과를 분석하고, 5장에서는 결론을 제시한다.

2. 추천 기법과 고객 프로파일링 기법

본 논문에서는 추천 기법의 성능을 향상시키기 위한 새로운 고객 프로파일링 기법을 제시하고자 한다. 이에 앞서 여기에서는 기존의 주요 추천 기법과 고객 프로파일링 기법에 대해 살펴보고자 한다.

2.1 추천 기법에 관한 연구

지금까지의 연구들에서 주로 논의되고 있는 추천 기법에는 내용기반 필터링(content-based filtering)과 협업 필터링(collaborative filtering)이 있다. 내용기반 필터링은 텍스트 집약(text intensive) 분야의 추천에 가장 효과적인 기법으로 [29], 과거 고객이 접근했던 데이터 아이템들(제품 또는 서비스)의 키워드, 구절, 특징을 파악하여 사용자의 관심을 나타내는 고객 프로파일로 정의하고, 아이템의 내용과 고객 프로파일 간의 유사성 비교를 통해 높은 관련성을 가지는 아이템들을 추천해 준다[27]. 내용기반 필터링을 적용한 추천 시스템에서는 선호된 아이템들에 대한 문서 정보를 분석하고, 유사한 정보를 가진 아이템들의 발견을 통하여 제품 추천이 이루어진다. 이 기법은 충분한 내용정보가 있을 경우에 적합하기 때문에 주로 기사 또는 웹 페이지와 같은 텍스트 정보의 추천에 유익하다. 내용기반 필터링을 적용한 추천시스템으로는 NewsWeeder[19], Letizia[23], Web-Watcher[15], InfoFinder[17]. 등이 있다. 그러나 내용기반 필터링은 텍스트 문서의 특징 추출에는 유용하나 동영상, 사진, 음악 등과 같은 몇몇 분야의 제품들에 대해서는 유용한 특징 추출이 불가능하다는 단점이 있다. 그리고 고객의 이전 경험과 유사한 제품만을 추천받을 수 있으며, 평가된 아이템들이 희소할 경우는 추천이 잘 이루어지지 못한다는 단점을 가지고 있다[6, 18, 20].

정보 검색 분야와 e-커머스 분야 모두에서 가장 성공적인 추천 기법으로 인식되고 있는 협업 필터링은 새로운 사용자의 알려지지 않은 선호를 예측하기 위하여 사용자들 그룹의 알려진 선호도를 사용하는 기법이다. 이 기법의 기본적인 가정은 만약 사용자 A와 B가 k 아이템들에 대해 유사하게 평가(rating)하였다면, 그들은 유사한 취향을 가지고 있을 것으로 판단되기 때문에 다른 아이템들에 대해서도 유사하게 평가한다는 것이다 [11]. 협업 필터링에서는 사용자들에게 아이템에

대한 직접적인 평가를 요구하고, 이러한 평가 정보를 바탕으로 유클리드 거리, 코사인 유사도, 상관관계 등과 같은 유사성 분석을 통해 최근접 이웃(nearest neighbor)을 발견해 낸다. 내용기반 필터링이 데이터 아이템과 사용자 프로파일 간의 유사성을 비교하기 때문에 사용자 개인의 내부에서 기인하는 추천이라고 한다면, 협업 필터링은 사용자 간의 유사성을 계산하기 때문에 사용자 외부의 다른 사람으로부터 오는 추천이라고 볼 수 있다[2]. 뿐만 아니라 이 기법은 내용기반 필터링과는 달리 아이템들에 대한 어떠한 분석도 고려하지 않으며, 음악, 이미지, 음성과 같은 영역에서도 활용 가능하다. 협업 필터링은 Goldberg 등[10]에 의해 개발된 Tapestry에서 처음으로 사용된 이래, Ringo[33], GroupLens[16] 등과 같은 다양한 추천시스템들에서 사용되고 있다.

그러나 협업 필터링은 자료의 희소성(sparsity)과 확장성(scalability)이라는 문제점을 가지고 있다[12]. 이러한 문제점으로 인해 협업 필터링에서는 사용자들에 의해 아이템들이 충분히 평가되기 전까지는 새로운 아이템이 추천될 수 없으며, 사용자들이 늘어날수록 처리해야 하는 자료의 양이 기하급수적으로 증가하여 처리에 많은 시간을 요하게 된다. 협업 필터링의 가장 큰 문제점인 희소성과 확장성을 해결하기 위해 분류(taxonomy)나 클러스터(cluster) 기법 등을 이용하여 분석할 상품이나 사용자의 수를 줄이고, 웹상에서의 고객 행동 정보 분석을 통해 제품 평가의 요구로 인한 고객의 번거로움과 이로 인한 평가 정보의 부족을 줄이고자 하는 연구가 수행되고 있다.

2.2 고객 프로파일링 기법에 관한 연구

정보과다의 문제를 해결하기 위한 좋은 대안인 개인화된 정보를 추천하기 위해서는 고객의 선호를 먼저 파악해야 한다. 웹 기술의 발전은 고객에 대한 가치 있는 정보를 보다 쉽게 얻을 수 있게 해주었고, 수집된 고객 정보의 분석을 통해 고객

선호도를 파악하여 고객이 원하는 정보를 추천해 줄 수 있게 하는 고객 프로파일의 정의를 가능하게 해주었다[35]. 고객 프로파일은 사용자 요구(needs)에 대한 구조화된 표현으로 내용기반 필터링에서는 고객 선호와 유사한 아이템의 추천을 위해 사용될 수 있고, 협업 필터링에서는 비슷한 취향이나 선호를 가진 고객의 그룹화 과정에 사용될 수 있다. Montaner 등[28]에 따르면, 정확한 고객 프로파일을 만드는 것은 시스템의 성공에 있어 핵심적인 사항일 뿐만 아니라 내용기반 필터링과 협업 필터링 모두에 있어 중요하다. 고객이 구매한 제품의 리스트와 고객의 평가 리스트를 사용하는 이력 기반 모델(history-based model)은 e-커머스에서 고객 프로파일을 정의하기 위해 가장 보편적으로 사용되는 접근법이다.

고객 프로파일링에 필요한 고객 정보를 수집하기 위한 방법은 연구자별로 다른 용어들로 사용되고 있다. Middleton 등[26]은 설문지와 인터뷰 등의 방법을 이용하여 고객 정보를 수집한 후, 고객에 가장 근접한 공학적 통계 모델을 만드는 방법을 지식기반(knowledge-based) 기법이라고 하였고, 고객의 행동으로부터 유용한 패턴을 발견하기 위해 기계 학습법(machine learning techniques)을 사용하는 방법을 행위기반(behavior-based) 기법이라고 하였다. Fan 등[9]은 정보 제품(information goods)의 사용자 프로파일링에 필요한 고객 정보 수집 방법으로 사용자에게 의해 직접적으로 제공되는 명시적 프로파일(explicit profile)과 사용자의 피드백과 행동 추적을 바탕으로한 푸시 시스템(push system)에 의해 만들어진 암묵적 프로파일(implicit profile)이 있다고 하였다. 비록 고객 프로파일링을 위한 고객 정보수집 방법은 연구자 별로 서로 다른 용어들로 사용되고 있지만, 이 기법들의 기본적인 개념은 고객으로부터 그들의 관심 분야에 대한 정보를 직접 입력 받는가? 아니면, 웹상에서의 고객 행동 정보의 분석 결과를 이용하여 간접적으로 고객의 관심 분야를 예측하는 것인가?라고 정리할 수 있다. 최근 대부분

의 추천시스템에서는 정보수집의 편리성과 데이터 분석 기술의 발전으로 웹상에서의 고객 행동 분석을 통해 고객 프로파일이 정의되고 있다.

행위기반 또는 암묵적인 방법을 통해 수집된 고객의 다양한 행동 정보 중 구매정보는 고객의 관심을 나타내는 주요 지표와[7] 고객 프로파일을 구축하기 위한 주요 정보로 활용되고 있다. 전통적인 협업 필터링에서는 추천의 첫 단계인 고객 프로파일을 구축하기 위해 고객의 구매기록을 이용하고 있으며[31, 32], Cho 등[8]은 추천 성능을 개선하기 위하여 시간의 흐름에 따라 고객의 구매 정보를 수집하고, 이 정보를 이용하여 동적인 고객 프로파일을 구축하였다. 그리고 Weng 등[34]은 제품 특징과 고객의 구매행동에 기반한 추천 기법을 제시하였다. 이 기법에서는 제품 특징들에 대하여 고객 선호도를 분석하여 고객 프로파일을 정의한 후, 제품 특징과 고객 프로파일 간의 유사성 분석을 통해 고객이 매력을 가질만한 제품을 추천해 준다. 이 기법은 한 번도 구매되지 않은 신제품 또는 구매가 거의 발생하지 않았거나 평가수가 희소한 제품들에 대해 낮은 추천률을 보이는 기존 추천 기법들의 단점을 개선하였다.

3. 고객 행동 기반 고객 프로파일링과 제품 추천 기법

여기에서는 고객 행동기반 고객 프로파일링 기법을 제시하고, 고객 프로파일과 제품 프로파일을 이용한 개인화된 추천 기법을 제시한다.

3.1 제품 프로파일

제품 프로파일(product profile)은 제품의 특징을 정리해 놓은 것으로, 고객 프로파일과의 유사성 계산을 통해 추천 제품을 결정할 때 이용된다. 제품 m 에 대한 제품 프로파일 PP_m 은 다음과 같이 정의된다[34].

$$PP_m = (f_m^j, i=1, \dots, I, \quad j=1, \dots, K_i) \\ m=1, \dots, M$$

여기서 I 는 제품특징의 수를 나타내고, K_i 는 제품특징 i 를 구성하는 특징값의 수를 나타낸다. M 은 제품의 수를 나타내고, f_m^j 는 제품 m 이 제품특징/특징값 i/j (제품특징 i , 특징값 j)를 가지고 있으면 1, 그렇지 않으면 0의 값을 가진다.

본 논문에서는 고객 행동에 기반한 개인화된 도서정보를 추천하기 위해 <표 1>과 같은 도서분류 체계를 이용하여 도서군에 대한 제품 프로파일을 정의한다. 즉, 제품특징(i)는 문학, 외국어, 종교, 실용서적, 컴퓨터/인터넷, 과학, 예술, 청소년, 어린이, 수험서로 정의되고(즉, $I=10$, $i=1$ (문학), $i=2$ (외국어), $i=3$ (종교), $i=4$ (실용서적), $i=5$ (컴퓨터/인터넷), $i=6$ (과학), $i=7$ (예술), $i=8$ (청소년), $i=9$ (어린이), $i=10$ (수험서)), 문학이라는 특징의 특징값(j)는 한국문학, 외국문학, 문학이론, 문학일반으로 정의된다(즉, $K_1=4$, $j=1$ (한국문학), $j=2$ (외국문학), $j=3$ (문학이론), $j=4$ (문학일반)). 문학 외의 나머지 특징의 특징값들도 동일한 방법으로 정의할 수 있다. 예를 들어 “문학/외국문학”이라는 제품특징/특징값을 가지는 도서군에 대한

제품 프로파일을 정의하면 다음과 같다.

$$PP_m = (f_m^{11}, f_m^{12}, f_m^{13}, f_m^{14}, f_m^{21}, \dots, f_m^{10,4}) \\ = (0, 1, 0, 0, 0, \dots, 0)$$

3.2 고객 행동 기반 고객 프로파일링 기법

고객 프로파일은 각 제품특징과 특징값에 대한 고객의 관심 수준을 정리해 놓은 것이다. 본 논문에서는 고객의 선호를 보다 잘 반영하는 고객 프로파일을 정의하기 위해 고객의 구매횟수 정보 이외에, 웹상에서 발생할 수 있는 고객의 다양한 직·간접적 관심표명 행동 중에서 고객의 선호와 가장 밀접한 관련이 있는 정보로 판단되는 고객의 클릭 횟수, 장바구니 담기 횟수, 관심 분야 정보를 이용하여 고객 프로파일을 구하였다.

고객 프로파일을 정의하기 위하여 먼저 고객 n 의 제품특징/특징값 i/j 에 대한 절대적 관심 수준(absolute customer interest) ACI_n^{ij} 를 다음과 같이 구한다.

$$ACI_n^{ij} = \alpha_1 \frac{b_n^{ij}}{B_n} + \alpha_2 \frac{c_n^{ij}}{C_n} + \alpha_3 \frac{s_n^{ij}}{S_n} + \alpha_4 \frac{p_n^{ij}}{P_n} \\ (\text{단, } \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1)$$

<표 1> 도서 분류 체계

구 분	항 목
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

여기서 B_n 은 고객 n 이 구매한 제품의 총 수를 나타내고, b_n^{ij} 는 고객 n 이 제품특징/특징값 i/j 를 가지는 제품을 구매한 횟수를 나타낸다. C_n 은 고객 n 이 클릭한 제품의 총 수를 나타내고, c_n^{ij} 는 고객 n 이 제품특징/특징값 i/j 를 가지는 제품을 클릭한 횟수를 나타낸다. S_n 은 고객 n 이 장바구니 담기를 한 제품의 총 수를 나타내고, s_n^{ij} 는 고객 n 이 제품특징/특징값 i/j 를 가지는 제품을 장바구니에 담은 횟수를 나타낸다. P_n 은 고객 n 이 관심 분야로 체크한 제품특징/특징값 i/j 의 총 수를 나타내고, p_n^{ij} 는 제품특징/특징값 i/j 를 가지는 제품에 대한 고객 n 의 관심 분야 체크 여부를 나타낸다. 이 때 고객 n 이 제품특징/특징값 i/j 를 관심 분야로 체크하였다면 1, 그렇지 않다면 0의 값을 부여하기로 한다. α 는 항목들 간의 상대적인 중요도를 반영하기 위한 가중치로 고정된 값을 사용할 수 있고, 각 행동 항목들의 중요도 변화를 고려하기 위해 조정할 수도 있다.

예를 들어 고객 n 의 총 구매 횟수가 8회, 총 클릭 횟수가 25회, 총 장바구니 담기 횟수가 12회, 회원가입시 관심 분야로 체크한 도서군의 수가 4개라고 가정하자. 그리고 고객 n 의 “문학/외국문학” 도서군에 속한 도서의 구매횟수가 2회, 클릭 횟수가 4회, 장바구니 담기 횟수가 3회이고, 회원가입시 이것을 관심 분야로 체크하였다고 가정하자. “문학/외국문학” 도서군에 대한 고객 n 의 ACI_n^{12} 는 $\alpha_1 \frac{2}{8} + \alpha_2 \frac{4}{25} + \alpha_3 \frac{3}{12} + \alpha_4 \frac{1}{4}$ 이 된다. 이때 모든 고객 행동 정보에 대하여 $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.25$ 로 동일하게 사용하기로 한다면, “문학/외국문학” 도서군에 대한 ACI_n^{12} 는 0.2275가 된다.

다음으로 고객 n 의 제품특징/특징값 i/j 에 대한 절대적 관심 수준인 ACI_n^{ij} 를 전체 고객에 대하여 상대적인 값으로 계산한 상대적 관심수준 RCI_n^{ij} (relative customer interest)는 다음과 같이 정의한다.

$$RCI_n^{ij} = \frac{ACI_n^{ij}}{\sum_{i=1}^N ACI_i^{ij}}$$

여기서 N 은 고객의 수를 나타낸다. 예를 들어 전체 고객들의 “문학/외국문학” 도서군에 대한 ACI_n^{12} 의 합이 1.24라고 한다면 고객 n 의 “문학/외국문학” 도서군에 대한 RCI_n^{12} 는 $0.1834 (= 0.2275 / 1.24)$ 가 된다.

ACI_n^{ij} 와 RCI_n^{ij} 가 계산되어지고 나면, 고객 n 에 대한 고객 프로파일 CP_n 을 다음과 같이 정의한다.

$$CP_n = (RCI_n^{ij}, i = 1, \dots, I \quad j = 1, \dots, K_i) \\ n = 1, \dots, N$$

고객 n 의 “문학/외국문학” 도서군에 대한 RCI_n^{12} 를 이용하여 CP_n 을 정의하면 $CP_n = (RCI_n^{11}, 0.1834, RCI_n^{13}, \dots, RCI_n^{10,1}, RCI_n^{10,2}, RCI_n^{10,3}, RCI_n^{10,4})$ 과 같다. 고객 n 의 나머지 도서군에 대한 ACI_n^{ij} , RCI_n^{ij} 의 계산 과정도 이와 동일하며, 이 과정을 통해 고객 프로파일이 완성된다.

3.3 제품 프로파일과 고객 프로파일을 이용한 제품 추천

유클리드 거리 계산법을 이용하여 고객 n 의 고객 프로파일 CP_n 과 각 제품 $m (m = 1, \dots, M)$ 의 제품 프로파일 PP_m 간의 유사성 R_{mn} 을 계산한다[34]. 계산된 유클리드 거리가 짧을수록 제품 프로파일과 고객 프로파일 간의 유사성이 높은 것을 의미하므로 고객이 선호하는 제품일수록 R_{mn} 의 값은 낮아진다.

$$R_{mn} = \sqrt{\sum_{ij} (f_m^{ij} - RCI_n^{ij})^2}, \quad m = 1, \dots, M$$

R_{mn} 값을 이용하여 제품을 추천하기 위해 추천하고자 하는 제품특징/특징값의 수들을 미리 정하고, R_{mn} 값이 작은 순서대로 제품특징/특징값을 갖는 제품을 추천한다.

<표 1>과 같은 제품특징/특징값을 가지는 도

서군 중에서 고객이 선호하는 도서군에 속한 도서를 추천하기 위해 유클리드 거리 계산법을 이용한 PP_m 과 CP_n 간의 유사성 계산 과정은 다음과 같다. 고객 n 의 $CP_n = (0, 0.1834, \dots, 0.561, 0)$ 이라고 한다면, “문학/한국문학” 도서군과의 R_{mn} 은

$$\sqrt{(1-0)^2 + (0-0.1834)^2 + \dots + (0-0.561)^2 + (0-0)^2}$$

$$= 1.1611$$

이 되고, “문학/외국문학” 도서군과의 R_{mn} 은

$$\sqrt{(0-0)^2 + (1-0.1834)^2 + \dots + (0-0.561)^2 + (0-0)^2}$$

$$= 0.9907$$

이 된다. 이와 같이 40개의 도서군에 대한 PP_m 과 고객 n 의 CP_n 간의 R_{mn} 을 계산하면, “수험서/공무원 수험서” 도서군과의 R_{mn} 이

$$\sqrt{(0-0)^2 + (0-0.1834)^2 + \dots + (1-0.561)^2 + (0-0)^2}$$

$$= 0.4758$$

로 가장 작은 값을 가지게 된다. R_{mn} 이 가장 작은 도서군에 해당되는 도서를 한 권 추천하기로 한다면, “수험서/공무원 수험서” 도서군에 속한 도서가 추천된다.

4. 실험 결과

4.1 실험 개요

제시된 고객 행동 기반 고객 프로파일링 기법의 효과를 평가하기 위해 IIS 6.0 환경에서 ASP (active server pages)를 개발 언어로 사용하여 Weng 등[34]이 제시한 고객 프로파일링 기법(WPT : weng profiling technique)을 적용한 도서정보 추천시스템과 본 논문에서 제시한 고객 행

동 기반 고객 프로파일링 기법(CBCPT : customer behavior based customer profiling technique)을 적용한 도서정보 추천시스템을 구축하였다. 특히 CBCPT는 각 고객 행동 정보들에 대한 가중치를 모두 동일하게 사용하는 경우($\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.25$)와 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심 분야 정보에 대한 가중치를 다르게 사용하는 경우($\alpha_1 = 0.4, \alpha_2 = 0.1, \alpha_3 = 0.2, \alpha_4 = 0.3$)로 나누어 실험을 실시하였다. WPT와 CBCPT는 동일한 방법으로 제품 프로파일을 정의하는 반면, 고객 프로파일을 정의하는 방법은 서로 다르다. 즉, WPT는 구매 횟수만을 고려하는 반면 CBCPT는 3.2에서 살펴본 바와 같이 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 관심 분야 정보 등을 종합적으로 고려하여 고객 프로파일을 정의한다. WPT와 CBCPT는 1회 추천시 서로 다른 도서군에 포함되어 있는 5권의 도서를 추천하도록 하였다.

한편, WPT와 CBCPT에 사용될 고객 행동 정보를 수집하기 위해 <표 1>의 도서 분류 체계에 따라 도서정보를 제공해 주는 일반 도서정보 시스템을 구축하였다. 이 시스템은 고객이 관심을 가질만한 도서정보를 추천해 주는 것이 아니라 단순히 클릭된 도서에 대한 정보를 제공해 준다. 이 시스템으로부터 수집된 고객 행동 정보들은 본 논문에서 제시하는 프로파일링 기법의 효과를 평가하기 위해 WPT와 CBCPT에 동일하게 적용된다. 일반 도서정보 시스템에 대한 고객 행동 정보를 수집하기 위해 인터넷 사용이 익숙한 74명의 대학생 및 일반인들을 피실험자로 선정하고, 그 시스템을 이용하여 각각 10회씩 도서를 구입하도록 하였다. 그리고 시스템의 이용에 앞서 40개의 제품특징/특징값들로 구성되는 도서군에 대한 구체적인 선호의 정도를 1~5점까지의 점수를 이용하여 직접 평가하도록 하였다. 이와 같은 과정을 통해 최종 58명의 피실험자들에 대한 행동 정보와 각 도서군에 대한 평가 정보가 수집되었다.

4.2 실험 결과 분석

WPT와 CBCPT의 성능 평가는 예측 평가(prediction evaluation)와 추천 평가(recommendation evaluation) 측면에서 이루어졌다. 먼저 예측 평가는 개별 아이템의 예측성을 평가하는 것으로, 본 논문에서는 고객이 직접 제시한 선호도와 추천시스템이 예측한 선호도 간에 절대값 평균을 의미하는 MAE(mean absolute error)를 이용하여 통계적 정확도(statistical accuracy)를 평가하였다. MAE 값이 낮을수록 추천시스템이 예측한 선호도가 고객이 직접 제시한 선호도에 가까워진다고 할 수 있다[30]. MAE는 다음과 같이 정의된다.

$$MAE = \frac{1}{N \times C} \sum_{n=1}^N \sum_{i,j} |r_n^{ij} - p_n^{ij}|$$

r_n^{ij} : 고객 n 이 제품특징/특징값 i/j 에 대해 직접 입력한 선호도

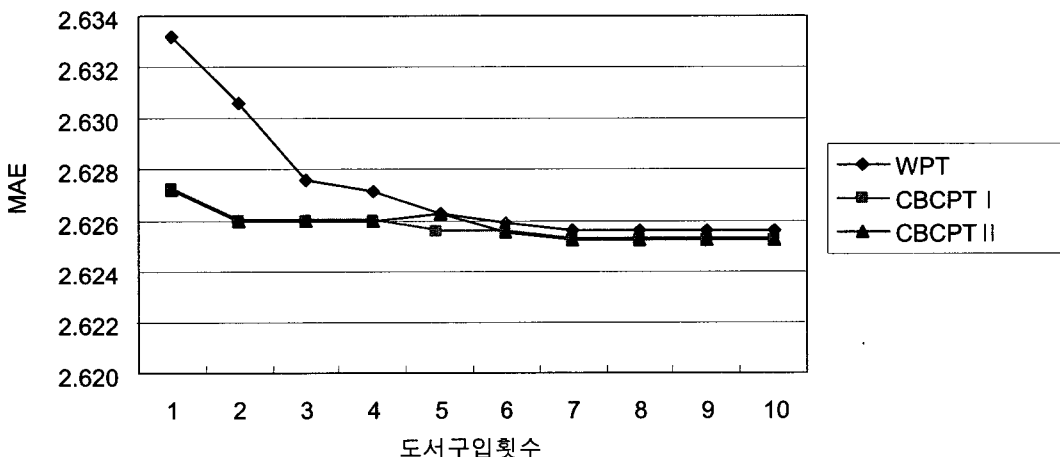
p_n^{ij} : 추천시스템이 제품특징/특징값 i/j 에 대해 예측한 고객 n 의 선호도

C : 고객이 평가한 제품특징/특징값의 총 수

총 58명의 피실험자들이 40개의 도서군에 대해

1~5점까지 평가하였고, WPT와 CBCPT에서는 40개의 도서군에 대한 추천 점수를 예측하므로 피실험자 1인당 40개의 평가값 - 예측값 쌍이 존재한다. 피실험자들로부터 직접 수집한 평가값과 도서 정보 추천시스템의 예측값을 이용하여 구한 MAE 값의 변화를 살펴보면 <그림 1>과 같다. 그림에서 도서구입 횟수는 1인당 구입횟수를 의미하고, CBCPT I은 $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.25$ 를 사용한 경우이고, CBCPT II는 $\alpha_1 = 0.4, \alpha_2 = 0.1, \alpha_3 = 0.2, \alpha_4 = 0.3$ 을 사용한 경우이다. 그림에서 보는 바와 같이 CBCPT I과 CBCPT II의 MAE는 거의 비슷한 것으로 나타났으며, 전체적으로 CBCPT의 MAE값은 WPT보다 작은 것으로 나타나 CBCPT가 WPT에 비해 피실험자들의 직접적인 평가값에 더 가까운 예측을 한다는 것을 알 수 있다. 뿐만 아니라 CBCPT의 MAE 값은 구입횟수가 적을(1회~4회) 경우에도 WPT에 비해 비교적 안정적이고 낮은 양상을 보이고 있다. 이를 통해 CBCPT는 실제 고객 행동에 대한 정보가 부족할 때에도 고객의 선호를 잘 반영하는 프로파일링 기법인 것을 알 수 있다.

다음으로 추천평가를 위해 정보검색 시스템 평가 지표로 가장 잘 알려져 있고, Basu 등[4], Billus 등[5], 그리고 Sarwar 등[16]에 의해 추천 시스템의 평가 지표로도 사용되어 오고 있는 정



<그림 1> WPT와 CBCPT의 MAE 비교

확률(precision)과 재현률(recall), 그리고 $F1$ 을 이용하였다. 정확률은 추천된 도서들이 고객이 선호하는 도서군의 도서일 확률을 의미하고, 재현률은 고객이 선호하는 도서군의 도서들이 추천되어질 확률을 의미한다. 여기서 고객이 선호하는 도서군이란 피실험자들에 의해 직접 제시된 평가값이 4 또는 5점인 도서군을 의미한다[30]. 그리고 $F1$ 은 추천된 제품의 수가 증가함에 따라 재현률은 증가하는 반면, 정확률이 감소하는 문제점을 해결하기 위해 이 두 지표에 동일한 가중치를 두어 하나의 평가항목으로 결합한 것이다. 정확률 P , 재현률 R , 그리고 $F1$ 은 다음과 같이 정의된다[14, 25].

$$P = \frac{N_{rs}}{N_s}$$

$$R = \frac{N_{rs}}{N_r}$$

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

N_s : 추천된 도서의 총 수

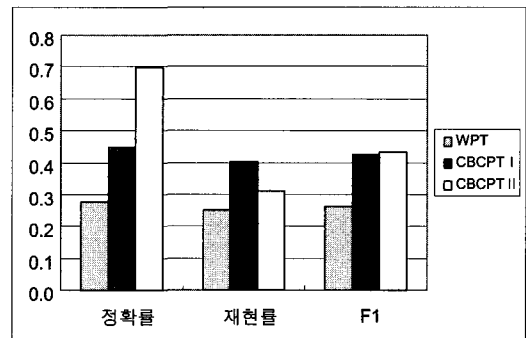
N_r : 선호하는 도서군의 총 수

N_{rs} : 추천된 도서가 선호하는 도서군의 도서인 수

WPT와 CBCPT I, II에서의 정확률, 재현률, $F1$ 을 비교한 결과가 <표 2>와 <그림 2>에 정리되어 있다. 표와 그림에서 보는 바와같이 CBCPT의 정확률, 재현률, $F1$ 모두가 WPT보다 높은 것으로 나타나 CBCPT는 WPT에 비해 더 좋은 추천 성능을 가지고 있는 것을 알 수 있다. CBCPT I은 CBCPT II에 비해 정확률은 낮은 반면 재현률은 높은 것으로 나타났고, 이 두 지표를 하나의 평가항목으로 결합한 $F1$ 은 약간 차이가 있는 것으로 나타났다. CBCPT I과 CBCPT II의 결과에서 보듯이 가중치에 따라 정확률, 재현률, $F1$ 의 차이가 발생할 수 있다. 따라서 추천 기법의 성능

<표 2> WPT와 CBCPT의 정확률, 재현률, $F1$ 비교

		정확률	재현률	$F1$
WPT		0.278	0.249	0.263
CBCPT	CBCPT I	0.449	0.401	0.424
	CBCPT II	0.697	0.312	0.431



<그림 2> WPT와 CBCPT의 정확률, 재현률, $F1$ 비교

을 보다 높일 수 있는 가중치를 선정하기 위한 연구가 추가적으로 필요할 것으로 보인다.

예측 평가와 추천 평가를 실시한 결과 CBCPT는 MAE, 정확률, 재현률, 그리고 $F1$ 에서 WPT보다 더 좋은 결과를 보이는 것으로 나타났다. 이를 통해 본 논문에서 제시한 고객 행동 기반 고객 프로파일링 기법인 CBCPT가 WPT에 비해 더 효과적인 고객 프로파일링 기법이라는 것을 알 수 있다.

5. 결 론

고객에게 개인화된 정보를 제공하기 위해서는 우선 고객이 어떠한 정보를 선호하는가를 나타내는 고객 프로파일을 정의하는 과정이 필요하다. 본 논문은 웹상에서의 다양한 고객 행동정보 중 고객의 선호와 가장 밀접한 관련이 있는 정보로 판단되는 구매횟수, 클릭 횟수, 장바구니 담기 횟수, 관심 분야 정보를 활용하여 보다 정확하게 고

객 프로파일을 정의하는 기법을 개발하였다.

이 기법은 제품의 특징 정보와 고객 행동정보에 기반하고 있다. 이전의 연구가 고객 프로파일을 작성하기 위해 구매 정보만을 이용하고 있는데 반해 본 논문에서는 웹상에서의 다양한 고객 행동 정보와 그 정보들의 실제 빈도를 사용하여 제품특징/특징값들에 대한 상대적인 관심 수준을 도출해 낸 후, 이것을 제품특징/특징값 별로 정리한 고객 행동기반 고객 프로파일을 정의하였다.

제시한 고객 행동기반 고객 프로파일링 기법의 효과를 분석하기 위해 Weng 등이 제시한 고객 프로파일링 기법 WPT를 적용한 도서정보 추천 시스템과 고객 행동기반 고객 프로파일링 기법 CBCPT를 적용한 도서정보 추천시스템을 구축하였다. 추천시스템의 예측 능력을 평가하기 위해 MAE를 사용하였고, 추천 능력을 평가하기 위해 정확률, 재현률, 그리고 $F1$ 을 사용하였다. 그 결과 CBCPT가 WPT에 비해 모든 평가 지표에서 우수한 것으로 나타났다.

본 논문은 고객의 선호를 나타내는 고객 프로파일을 정의할 때 구매 횟수, 클릭 횟수, 장바구니 담기 횟수, 그리고 관심 분야 정보와 같은 고객의 선호와 관련 있는 다양한 정보들의 실제 빈도를 활용한 새로운 고객 프로파일 모델을 제시하였다는데 그 의의가 있다. 본 논문에서 제시한 고객 프로파일 모델을 2장에서 살펴보았던 다양한 추천 기법에 활용한다면 고객이 원하거나 또는 관심을 가지고 있는 제품을 보다 정확하게 추천해 줄 수 있을 것이다. 이것은 제품정보 검색에 소비되는 고객의 시간과 노력을 줄여줄 것이고, 나아가 해당 사이트에 대한 고객의 만족도와 충성도를 향상시키게 될 것이다.

마지막으로, 향후에 아래와 같은 연구가 추가적으로 필요할 것으로 보인다. 첫째, 본 논문에서는 고객 개인 정보만을 이용하여 고객 프로파일을 구성하고 있기 때문에 고객이 현재 관심을 가지는 제품에 대한 추천은 비교적 잘 이루어지고 고객의 잠재된 구매 욕구를 자극할 수 있는 새로운

제품의 추천은 잘 이루어지지 못한다. 따라서 고객의 잠재된 구매 욕구를 자극할 수 있는 새로운 제품을 추천하기 위해 유사한 성향이나 특징을 가진 고객 그룹의 관심 표명 행동에 대한 정보를 고객 프로파일에 반영하는 연구가 필요할 것으로 보인다. 둘째, 고객의 다양한 관심 표명 행동들의 가중치에 따라 추천 기법의 성능이 달라질 수 있다. 따라서 고객의 다양한 관심 표명 행동들의 상대적인 중요도를 가장 잘 반영하는 최적 가중치 선정에 관한 연구가 필요할 것으로 보인다.

참 고 문 헌

- [1] 김종우, 배세진, 이홍주, "협업 필터링 기반 개인화 추천에서의 평가자료의 희소 정도의 영향", 「경영정보학연구」, 제14권, 제2호(2004), pp.131-149.
- [2] 이수정, 이형동, 김형주, "사용자 경향에 기반한 동적 추천 기법: 영화 추천 시스템을 중심으로", 「정보과학회논문지: 소프트웨어 및 응용」, 제31권, 제2호(2004), pp.153-163.
- [3] Ansari, A., S. Essegai, and R. Kohli, "Internet Recommendation Systems," *Journal of Marketing Research*, Vol.37, No.3(2000), pp.363-375.
- [4] Basu, C., H. Hirsh, and W. Cohen, "Recommendation as classification: using social and content-based information in recommendation," *Proceedings of the 1998 workshop on recommender systems*, (1998), pp.714-720.
- [5] Billsus, D. and M.J. Pazzani, "Learning Collaborative Information Filters," *Proceedings of ICML*, (1998), pp.46-53.
- [6] Changchien, S.W., C.F. Lee, and Y.J. Hus, "On-line personalized sales promotion in electronic commerce," *Expert Systems with Application*, Vol.27, No.1(2004), pp.35-52.

- [7] Cheung, K.W., J.T. Kwok, M.H. Law, and K.C. Tsui, "Mining customer product ratings for personalized marketing," *Decision Support Systems*, Vol.35, No.2(2003), pp.231-243.
- [8] Cho, Y.B., Y.H. Cho, and S.H. Kim, "Mining changes in customer buying behavior for collaborative recommendations," *Expert Systems with Applications*, Vol.28, No.2 (2005), pp.359-369.
- [9] Fan, W., M.D. Gordon, and P. Pathak, "Effective profiling of consumer information retrieval needs: a unified framework and empirical comparison," *Decision Support Systems*, Vol.40, No.2(2005), pp.213-233.
- [10] 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), pp.61-70.
- [11] Glodberg, K., T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: A Constant Time Collaborative Filtering Algorithm," *Information Retrieval Journal*, Vol.4, No.2(2001), pp.133-151.
- [12] Good, N., J.B. Schafer., J.A. Konstan., A. Borchers., B. Sarwa., J. Herlocker, and J. Riedl, "Combing collaborative filtering with personal agents for better recommendations," *Proceedings of the 16th National Conference on Artificial Intelligence*, (1999), pp.439-446.
- [13] Hayes, C., P. Cunningham, and B. Smyth, "A case-based reasoning view of automated collaborative filtering," *Proceedings of the 4th international conference on case-based reasoning*, (2001), pp.243-248.
- [14] Herlocker, J.L., J.A. Konstan, L.G. Terveen, and J.T. Riedl, "Evaluation Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, Vol.22, No.1(2004), pp.5-53
- [15] Joachims, T., D. Freitag, and T. Mitchell, "Webwatcher: A tour guide for the world wide web," *15th International Joint Conference on Artificial Intelligence*, (1997), pp.770-777.
- [16] Konsta, J., B. Miller, D. Maltz., J. Herlocker., L. Gordon, and J. Riedl, "GroupLens: Applying collaborative filtering to usenet news," *Communications of the ACM*, Vol.40, No.3(1997), pp.77-87.
- [17] Krulwich, B. and C. Burkey, "The Info-Finder agent : learning user interests through heuristic phrase extraction," *IEEE Expert*, Vol.12, No.5(1997), pp.22-27.
- [18] Kwak, M. and D.S. Cho, "Collaborative filtering with automatic rating for recommendation," *Proceedings of ISIE 2001 IEEE international symposium on industrial electronics*, (2001), pp.625-628.
- [19] Lang, K., "Newsweeder : Learning to filter netnews," *Proceedings of the 12th International Conference on Machine Learning*, (1995), pp.331-339.
- [20] Lee, C.H., Y.H. Kim, and P.K. Rhee, "Web personalization expert with combing collaborative filtering and association rule mining technique," *Expert Systems with Applications*, Vol.21, No.3(2001), pp.131-137.
- [21] Lee, W.P., "Applying domain knowledge and social information to product analysis and recommendations: an agent-based decision support system," *Expert Systems*, Vol.21, No.3(2004), pp.138-148.
- [22] Lee, W.P. and T.H. Yang, "Personalizing in-

- formation appliances : a multi-agent framework for TV programme recommendations," *Expert Systems with Applications*, Vol.25, No.3(2003), pp.331-341.
- [23] Leiberman, H., "Letizia : An agent that assist web browsing," *Proceeding of the International Joint Conference on Artificial Intelligence*, (1995), pp.924-929.
- [24] Li, Y., L. Lu, and L. Xuefeng, "A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce," *Expert Systems with Applications*, Vol.28, No.1(2005), pp.67-77.
- [25] Liu, D.R., and Y.Y. Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management*, Vol.42, No.3 (2005), pp. 387-400.
- [26] Middleton, S.E., N.R. Shadbolt, and D.C. de Roure, "Ontological User Profiling in Recommender Systems," *ACM Transactions on Information Systems*, Vol.22, No.1(2004), pp.54-88.
- [27] Min, S.H. and I. Han, "Detection of the customer time-variant pattern for improving recommender systems," *Expert Systems with Application*, Vol.28, No.2(2005), pp.189-199.
- [28] Montaner, M., B. Lopez, and J.L. De La Rosa, "A Taxonomy of Recommender Agents on the Internet," *Artificial Intelligence Review*, Vol.19, No.4(2003), pp.285-330.
- [29] Perugini, S., M.A. Goncalves, and E.A. Fox, "Recommender Systems Research : A Connection-Centric Survey," *Journal of Intelligent Information Systems*, Vol.23, No.2(2004), pp.107-143.
- [30] 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 Collaborative Filtering System," *Proceedings of CSCW 98*, (1998), pp.345-354.
- [31] Sarwar, B.M. G. Karypis., J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for E-commerce," *Proceedings of ACM E-commerce 2000 conference*, (2000), pp.158-167.
- [32] Sarwar, B.M., G. Karypis., J.A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithm," *Proceedings of the 10th international world wide web conference*, (2001), pp.285-295.
- [33] Shardanand, U. and P. Maes, "Social information filtering: Algorithms for automating word of mouth," *ACM Conference on Computer Human Interaction*, (1995), pp.210-217.
- [34] Weng, S.S. and M.J. Liu, "Feature-based recommendations for one-to-one marketing," *Expert Systems with Applications*, Vol.26, No.4(2004), pp.493-508.
- [35] Yen, B.P.C. and R.C.W. Kong, "Personalization of information access for electronic catalogs on the web," *Electronic Commerce Research and Applications*, Vol.1, No.1(2002), pp.20-40.