

개인화된 제품 추천을 위한 고객 행동 기반 고객 프로파일 모델 연구

A study on the customer behavior based customer profile model for personalized products recommendation

박유진¹ 장근녕²,

¹ 연세대학교 경영학과 박사과정 parkyoujin@yonsei.ac.kr

² 연세대학교 정경대학 경영학부 knchang@dragon.yonsei.ac.kr

Abstract

In this paper, we propose a new customer profile model based on customer behavior in Internet shopping mall. The proposed technique defines customer profile model based on customer behavior information such as click data, buy data, and interest categories. We also implement CBCPM(Customer Behavior-based Customer Profile Model) and perform extensive experiments. The experimental results show that CBCPM has higher precision, recall, and F1 than the existing customer profile model.

I. 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[18]. 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 제

품들을 쉽게 발견할 수 있도록 도움을 주는 것에 데이터 분석 기법을 적용한 시스템이다[4]. 현재 추천 시스템은 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[1],[3]. 추천시스템의 사용은 추가적인 제품 제시를 통해 교차판매(cross-sell) 가능성과 고객 충성도(loyalty)를 향상시키고, 고객이 구매하기를 원하는 제품 발견을 도와줌으로써 고객 욕구를 만족시켜준다.

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 고객이 어떠한 제품 또는 제품 영역에 관심이 있는가를 나타내는 고객 프로파일을 정의하는 과정이 필요하다. Lee et al.[21]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 연구

의 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 추천 시스템의 기법과 고객 프로파일 기법에 관한 기존 연구를 살펴보고, 3장에서는 고객의 행동 정보에 기반한 새로운 고객 프로파일 모델을 제시한다. 4장에서는 제시된 모델을 평가하기 위한 실험 및 실험 결과를 제시하고, 5장에서는 결론을 제시한다.

II. 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[17]. 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[21] 사용자의 프로파일은 설문지, 평가된 항목, 또는 사용자 항해 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[7],[14].

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.[9]에 의해 개발된 Tapestry에서 협업필터링 기법이 처음으로 사용된 이래, Lotus Notes, GroupLens, Firefly, SiteSeer, PHOAKS, Grassroots 등과 같은 다양한 추천시스템들에서 협업필터링이 사용되고 있다.

그러나 협업필터링은 자료의 희소성(sparsity)과 자료의 확장성(scalability)이라는 문제점을 가지고 있다[15]. 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 사용자에게 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.

web usage mining의 기본적인 절차는 데이터 수집(data collection), 데이터 전처리(data preprocessing), 패턴발견(pattern discovery), 지식 사후처리(knowledge post-processing)로 나누어진다. 데이터 수집 단계는 분석되어질 웹 데이터를 서버 로그파일(server log files), 쿠키(cookies), 사용자, 외부, 클라이언트 데이터(client data)로부터 수집하는 단계이다. 데이터 전처리 단계는 수집된 웹 데이터를 일관성 있고, 통합적이고, 포괄적인 관점으로

정렬하는 단계이고, 패턴발견 단계는 기계학습(machine learning)과 통계적 방법을 사용하여 전처리된 웹 데이터로부터 행동의 패턴을 발견하는 단계이다. 마지막 단계인 지식 사후처리 단계는 추출된 지식을 평가하고 인간이 이해할 수 있는 형태로 표현해 주는 단계이다[10].

web usage mining에서는 특히 클러스터링(clustering), 분류(classification), 연관규칙 발견(discovery of associations), 순차적 패턴 발견(sequential pattern discovery)과 같이 패턴발견 단계에 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 베이저안 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 프로파일링(user profiling) 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[16].

지식기반 기법은 고객으로부터 직접 데이터를 입력받기 때문에 명시적 프로파일(explicit profile)이라고도 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[19].

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[20]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.[22]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. Balabanovic et al.[12]에 의해 개발된 추천시스템인 Fab과 Pazzani et al.[13]에 의해 개발된 Syskill과 Webert는 Salton et al.[11]에 의해 제시된 TF-IDF(Term Frequency-Inverse Document Frequency)를 통해 고객이 현재 보고 있는 문서의 특징 추출을 통하여 고객의 관심을 계산하였다.

Weng et al.[18]은 고객의 구매 정보를 이진(binary) 정보로 표현하고 이들의 합과 고객이 구매한 제품 수와의 관계로 고객 프로파일을 작성하였다. 그런데 이 연구에서는 0과 1이라는 이진 체계로 표현된 구매 정보만을 고객 프로파일 모델 생성에 반영하고 있는데, 구매가 발생한 제품만을 고객이 선호한다고는 볼 수 없다. 왜냐하면 제품에 대한 클릭, 방문, 북마크, 인쇄, 그리고 해당 제품에 대한 고객의 직접적인 관심 여부 등과 같은 다양한 고객의 행동들도 선호에 영향을 미치기 때문이다. 그리고 구매 정보의 이진 표현은 단순히 그러한 행동이 일어났다는 것에 대한 정보만을 나타낼 뿐 구체적인 강도에 대한 정보는 포함하고 있지 못하다.

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.

III. 고객 행동 기반 고객 프로파일 모델

여기에서는 구매에 영향을 미치는 다양한 고객 행동 정보 중에서 고객 선호와 가장 밀접한 관련이 있는 정보로 판단되는 고객의 구매 횟수, 클릭 횟수, 관심분야 체크여부와 제품의 특징정보에 기반한 고객 프로파일 모델을 제시한다.

3.1 제품 프로파일 모델

제품 프로파일(product profile) 모델은 제품의 특성을 정리해 놓은 것으로, 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 i 를 구성하는 특징값의 수 ($i=1, \dots, I$)
- f_m^{ij} : 제품 m 이 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, $I=3, i=1$ (색상), $i=2$ (재질), $i=3$ (용도)), 색상이라는 특징의 구체적 특징값으로는 검정, 노랑, 파랑(즉, $K_1=3, j=1$ (검정), $j=2$ (노랑), $j=3$ (파랑)), 재질이라는 특징의 구체적인 특징값으로는 가죽과 천(즉, $K_2=2, j=1$ (가죽), $j=2$ (천)), 용도라는 특징의 구체적인 특징값으로는 서류용과 여행용(즉, $K_3=2, j=1$ (서류용), $j=2$ (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 고객 프로파일 모델

고객 프로파일 모델은 고객의 구매 횟수, 클릭 횟수, 관심분야 체크여부를 이용하여 각 제품 특징에 대한 고객의 관심 수준을 분석한 것으로, 우선 특정 고객A의 관심 수준을 계산해 내고, 이를 전체 고객에 대한 상대적인 수준으로 계산하여 만들어진

다.

고객A의 제품 특징 i, j (제품 특징 i , 특징값 j)에 대한 가중 관심 수준을 나타내는 WCI(Weighted Customer Interest)는 다음과 같이 구해진다.

$$WCI_A^{ij} = a_1 \frac{b_A^{ij}}{B_A} + a_2 \frac{c_A^{ij}}{C_A} + a_3 \frac{p_A^{ij}}{P_A}$$

- WCI_A^{ij} : 고객A의 제품 특징 i, j 에 대한 가중 관심
- B_A : 고객 A가 구매한 제품의 총 수
- b_A^{ij} : 제품 특징 i, j 를 만족하는 제품에 대한 고객A의 구매 횟수
- C_A : 고객 A가 클릭한 제품의 총 수
- c_A^{ij} : 제품 특징 i, j 를 만족하는 제품에 대한 고객A의 클릭 횟수
- P_A : 고객 A가 관심분야로 체크한 제품의 총 수
- p_A^{ij} : 제품 특징 i, j 를 만족하는 제품에 대한 고객A의 관심분야 체크 여부
- a_1 : 구매에 대한 가중치
- a_2 : 클릭에 대한 가중치
- a_3 : 관심분야 체크에 대한 가중치 ($a_1 + a_2 + a_3 = 1$)

WCRI(Weighted Customer Relative Interest)는 제품 특징 i, j 에 대한 고객A의 관심 수준을 전체 고객과 비교하여 상대적으로 계산한 값으로 다음과 같이 구해진다.

$$WCRI_A^{ij} = \frac{WCI_A^{ij}}{\frac{1}{|S|} \sum_{s \in S} WCI_s^{ij}}$$

S : 모든 고객의 집합

고객A의 고객 프로파일 CP(Customer Profile)는 $WCRI_A^{ij}$ 를 사용하여 다음과 같이 표현되어진다.

$$CP_A = (WCRI_A^{ij}, i=1, \dots, I, j=1, \dots, K_j)$$

이상에서 보는 바와 같이 본 연구에서 제시하는 고객 프로파일 모델은 각 제품 특징 i, j 에 대한 고객의 클릭 횟수와 구매 횟수, 그리고 관심분야 체크 여부를 반영한 고객의 상대적인 관심의 수준이다.

3.3 제품 추천 모델

여기에서는 앞에서 제시한 제품 프로파일 모델과 고객 프로파일 모델을 이용한 개선된 제품추천

모델을 제시한다. 먼저 개인화된 정보를 추천하기 위해 고객과 제품 간의 유사성을 계산한다. 고객A와 제품 m간의 유사성은 유클리드 거리 계산법을 이용하여 다음과 같이 계산되어진다.

$$R_{Am} = \sqrt{\sum_{ij} (WCRI_A^{ij} - f_m^{ij})^2}$$

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

IV. 실험 및 결과 분석

4.1 실험 데이터

추천시스템의 성능 평가를 위해 실제 운영되고 있는 상업용 인터넷 쇼핑몰의 데이터와 MovieLens¹⁾나 EachMovie²⁾와 같이 일부 연구소에서 제공하고 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 Server 2003, 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.이 제시한 고객 프로파일 모델을 적용한 도서정보 추천시스템과 클릭 횟수, 구매 횟수, 관심분야 체크 여부와 같은 고객 행동 정보를 반영하는 고객 행동 기반 고객 프로파일 모델(CBCPM: Customer Behavior-based Customer Profile Model)을 적용한 도서정보 추천시스템을 구축하였다.

구축된 도서정보 추천시스템들의 피실험자로 인터넷 사용이 익숙한 74명의 대학생 및 일반인을 선

정하였으며, 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, 그리고 관심분야 정보가 수집되었으며, 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.[8], Basu et al.[6], Sarwar et al.[4] 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.

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

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

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

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 협업필터링 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.

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

추천되는 도서 정보의 크기가 1에서 5일 때 까지의 시스템별 평균 정확률과 재현률, 그리고 F1을 비교하였다.

Weng 모델과 CBCPM을 비교한 결과가 <표 2>와 [그림 1]~[그림 3]에 정리되어 있다. 표와 그림에서 CBCPM의 가중치 $\alpha_1=0.4$, $\alpha_2=0.2$, $\alpha_3=0.4$ 로 가정하였다.

Weng 모델과 CBCPM의 정확률을 비교해 놓은 <표 2>와 [그림 1]에서 보는 바와 같이 모든 추천 크기에서 CBCPM의 정확률이 Weng 모델보다 높게 나타났다. 이것은 고객 프로파일을 작성하기 위해 이진 체계로 표현된 고객의 구매 정보만을 반영할 때보다는 구매 정보와 더불어 클릭 횟수와 관심분야 체크여부에 대한 정보를 반영할 때, 추천 도서 중에서 고객이 실제 선호하는 도서의 수를 나타내는 정확률이 증

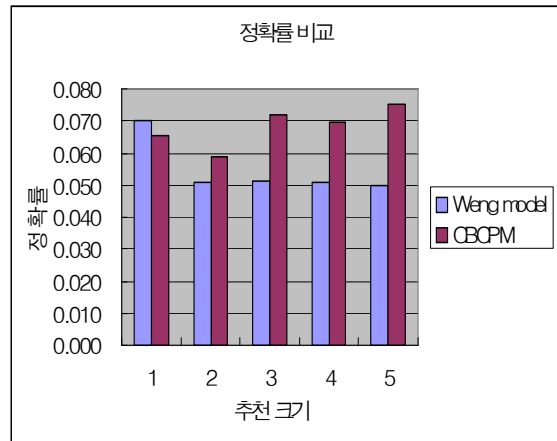
<표 2> 시스템 별 평균 정확률, 재현률, F1

추천크기 \ 측정값	Weng 모델		
	정확률	재현률	F1
1	0.070	0.070	0.070
2	0.051	0.094	0.065
3	0.051	0.122	0.070
4	0.051	0.138	0.072
5	0.050	0.138	0.070
추천크기 \ 측정값	CBCPM		
	정확률	재현률	F1
1	0.066	0.066	0.066
2	0.059	0.117	0.078
3	0.072	0.210	0.107
4	0.070	0.248	0.108
5	0.075	0.314	0.120

가한다는 것을 의미한다.

Weng 모델과 CBCPM의 재현률을 비교해 놓은 <표 2>와 [그림 2]에서 보는 바와 같이 모든 추천 크기에서 CBCPM의 재현률이 Weng 모델보다 높게 나타났다. 이것은 고객 프로파일을 작성하기 위해 이진 체계로 표현된 고객의 구매 정보만을 반영할 때보다

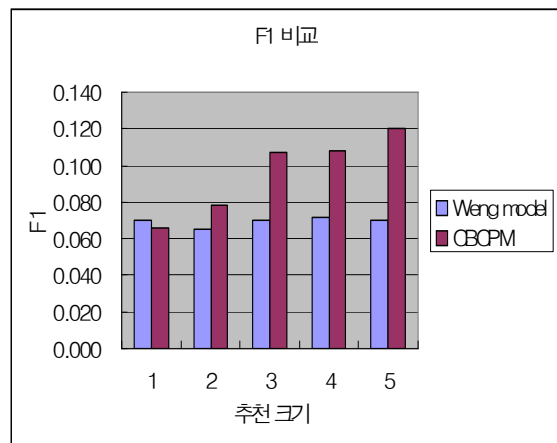
[그림 1] Weng 모델과 CBCPM의 정확률 비교



는 구매 정보와 더불어 클릭 횟수와 관심분야 체크여부에 대한 정보를 반영할 때, 고객이 선호하는 도서가 실제 추천되는 수를 나타내는 재현률이 증가한다는 것을 의미한다.

Weng 모델과 CBCPM의 F1을 비교해 놓은 <표 2>와 [그림 3]에서 보는 바와 같이 모든 추천 크기에서 CBCPM의 F1이 Weng 모델보다 높게 나타났다. 이것은 고객 프로파일을 작성하기 위해 이진 체계로 표현된 고객의 구매 정보만을 반영할 때보다는 구매 정보와 더불어 클릭 횟수와 관심분야 체크여부에 대한 정보를 반영할 때 정확률과 재현률을 하나의 평가 항목으로 결합한 F1이 더 증가한다는 것을 의미한다.

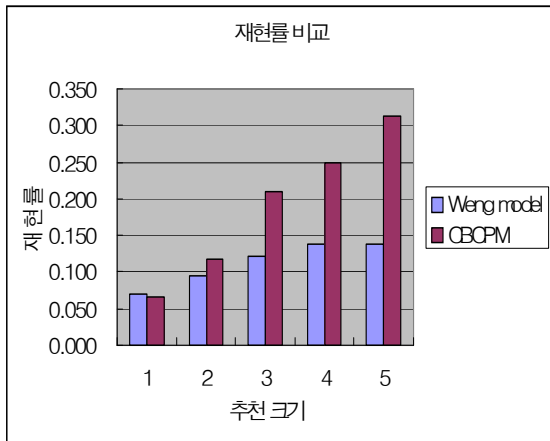
[그림 2] Weng 모델과 CBCPM의 F1 비교



V. 결론

고객에게 개인화된 정보를 제공하기 위해서는 우선 고객이 어떠한 정보를 선호하는가를 나타내는 고객 프로파일을 정의하는 과정이 필요하다. 본 연구는 구매에 영향을 미치는 여러 가지 구매행동 중

[그림 3] Weng 모델과 CBCPM의 재현률 비교



구매와 고객의 관심 수준에 가장 밀접한 관련이 있는 정보로 판단되는 고객의 구매 횟수, 클릭 횟수, 관심분야 체크여부에 관한 정보를 활용하여 보다 정확하게 고객 프로파일을 정의하는 기법을 개발하였다.

이 기법은 제품의 특징 정보와 고객 행동 정보에 기반하고 있다. 이전의 연구들이 고객 프로파일을 작성하기 위해 이진 정보 형태의 구매 정보만을 사용하고 있는데 반해 본 기법은 실제 발생한 구매 횟수와 클릭 횟수, 그리고 고객이 직접 입력한 관심분야 정보를 사용하였다. 이러한 정보를 토대로 제품 특징들에 대한 상대적인 가중 관심 수준을 도출해 낸 후 이것을 제품 특징별로 정리한 고객 행동 기반 고객 프로파일을 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.이 제시한 고객 프로파일 모델을 적용한 도서정보 추천시스템과 구매 횟수, 클릭 횟수, 그리고 관심분야 체크여부를 동시에 반영하는 고객 행동 기반 고객 프로파일 모델 CBCPM을 이용한 도서정보 추천시스템을 구축하였다. 각각의 시스템에 대하여 정확률, 재현률, 그리고 F1을 평가한 결과 CBCPM이 Weng 모델에 비해 모든 평가 지표에서 우수한 것으로 나타났다.

본 연구는 고객의 선호를 나타내는 고객 프로파일을 정의할 때 이진 체계로 표현된 구매 정보 대신 클릭 횟수, 구매 횟수, 그리고 관심분야 체크여부와 같은 실제적인 고객 행동 정보를 활용한 새로운 고객 프로파일을 제시하였다는 데 그 의의가 있다. 본

논문에서 제시한 고객 프로파일 모델을 2장에서 살펴 보았던 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 줄여 줄 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 연구는 고객 개인의 정보만을 이용하여 고객 프로파일을 구성하고 있기 때문에 고객이 구매 또는 클릭을 한 번도 하지 않았거나 또는 관심분야로 체크하지 않은 정보에 대한 추천은 제공하고 있지 못하다. 그리고 가중치 α 가 제품 특징에 대한 관심의 수준에 어떠한 영향을 미치는 지에 대한 다양한 분석 없이 이전 연구 경험을 토대로 미리 그 값을 지정해 놓았으며, 보다 다양한 고객 행동 정보를 고객 프로파일 모델에 반영하지 못하였다는 데 연구의 한계가 있다.

그러므로 협업 필터링이나 클러스터링 기법 등을 통해 고객과 유사한 관심 분야를 가지고 있는 그룹을 정의한 후, 그룹 구성원에 대한 정보도 고객 프로파일 형성에 반영도록 해야 할 것이다. 뿐만 아니라 분석을 통한 최적 가중치 α 의 선정과 다양한 고객 행동 정보를 반영하는 고객 프로파일 모델에 관한 연구도 추가적으로 이루어져야 할 것으로 사료된다.

<참고 문헌>

- [1] 김중우, 배세진, 이홍주, “협업 필터링 기반 개인화 추천에서의 평가자료의 희소 정도의 영향,” 한국경영정보학회, 경영정보학연구, 14(2), 2004, pp.131-149.
- [2] 이용준, 이세훈, 왕창중, “협업 여과 기반의 교육용 콘텐츠 추천시스템 설계,” 한국컴퓨터교육학회 논문지, 6(2), 2003, pp.147-156.
- [3] A. Ansari, S. Essegai, and R. Kohli, "Internet Recommendation Systems", Journal of Marketing Research, 37(3), 2000, pp. 363-375.
- [4] B. M. Sarwar, 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, Seattle, Washington, 1998, pp.345-354.
- [5] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Analysis of recommendation algorithms for E-commerce", Proceedings of the 2nd ACM Conference on Electronic Commerce(EC'00), ACM, New York, 2000, pp.285-295.
- [6] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: using social and content-based information in recommendation", Proceedings of the 1998 workshop on recommender systems, Menlo Park, CA: AAAI Press, 1998, pp.714-720.
- [7] C. H. Lee, Y. H. Kim, and P. K. Rhee, "Web personalization expert with combing collaborative filtering and association rule mining technique", Expert Systems with Applications, 21(3), 2001, pp.131-137.
- [8] D. Billsus, and M. J. Pazzani, "Learning Collaborative Information Filters", Proceedings of ICML, 1998, pp.46-53.
- [9] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry", Communications of the ACM, 35(12), 1992, pp.61-70.
- [10] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. D. Spyropoulos, "Web Usage Mining as a Tool for Personalization: A Survey", User Modeling and User-Adapted Interaction, 13(4), 2003, pp.311-372.
- [11] G. Salton, and M. J. McGill, "An Introduction to Modern Information Retrieval", McGraw-Hill, New York, 1983.
- [12] M. Balabanovic, and Y. Shoham, "Fab: Combing content-based and collaborative recommendation", Communications of the ACM, 40(3), 1997, pp.66-72.
- [13] M. J. Pazzani, J. Muramatsu, and D. Billsus, "Syskill & webert: Identifying interesting web sites", Proceeding of the 13th National Conference on Artificial Intelligence, 1996.
- [14] M. Kwak, 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.
- [15] N. Good, 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 Sixteenth National Conference on Artificial Intelligence, Orlando, FL, 1999, pp.439-446.
- [16] S. E. Middleton, N. R. Shadbolt, and D. C. de Roure, "Ontological User Profiling in Recommender Systems", ACM Transactions on Information Systems, 22(1), 2004, pp.54-88.
- [17] S. H. Min, and I. Han, "Detection of the customer time-variant pattern for improving recommender systems", Expert Systems with Application, 28(2), 2005, pp.189-199.
- [18] S. S. Weng, and M. J. Liu, "Feature-based recommendations for one-to-one marketing", Expert Systems with Applications, 26(4), 2004, pp.493-508.
- [19] W. Fan, M. D. Gordon, and P. Pathak, "Effective profiling of consumer information retrieval needs: a unified framework and empirical comparison", Decision Support Systems, 40(2), 2005, pp.213-233.
- [20] W. P. Lee, "Applying domain knowledge and social information to product analysis and recommendations: an agent-based decision support system", Expert Systems, 21(3), 2004, pp.138-148.
- [21] W. P. Lee, and T. H. Yang, "Personalizing information appliances: a multi-agent framework for TV programme recommendations". Expert Systems with Applications, 25(3), 2003, pp.331-341.
- [22] Y. B. Cho, Y. H. Cho, and S. H. Kim, "Mining changes in customer buying behavior for collaborative recommendations", Expert Systems with Applications, 28(2), 2005, pp.359-369.