

설명기능을 추가한 협업필터링 기반 개인별 상품추천시스템: WebCF-Exp

김재경

경희대학교 경영대학 e비즈니스 전공 교수
(jaek@khu.ac.kr)

이희애

경희대학교 e비즈니스 정보센터 연구원
(leeheeae@hanmail.net)

안도현

경희대학교 일반대학원 경영학과 박사과정
(adh@khu.ac.kr)

조윤희

국민대학교 경성대학 e비즈니스 학부 조교수
(www4u@kookmin.ac.kr)

본 연구는 기존의 협업필터링 기반 상품추천시스템의 추천품질과 신뢰도를 높이기 위하여 설명기능을 추가한 WebCF-Exp 추천시스템을 개발하고 그 시스템의 성능을 평가하고자 한다. 기존의 협업필터링 기반 상품추천시스템은 추천 성능의 우수함에도 불구하고 사용자들에게 추천의 이유를 설명하는 기능이 없으므로, 고객이 추천한 상품에 대한 신뢰감이 부족해지기 쉬운 단점이 있다. 이 연구는 기존의 웹마이닝과 상품계층도를 적용한 협업필터링 기반 상품추천 시스템에 추천한 이유를 고객에게 제공함으로써 고객이 그 시스템을 신뢰할 수 있도록 도움을 주기 위한 여러 형태의 설명기능을 제공한다. 또한 이러한 설명기능을 추가한 추천시스템 WebCF-Exp를 개발하고, 그 효용성을 검증하기 위한 실험을 위하여 인터넷 쇼핑몰을 구현하였다. 마지막으로, 가상의 인터넷 쇼핑몰을 이용한 고객들을 대상으로 온라인 설문을 통해 각기 다른 20개의 설명인터페이스 중 과연 어떤 인터페이스가 고객들이 이해하기에 가장 적합한지와 추천된 상품에 설명기능이 추가되었을 때 사용자들의 구매 의사결정에 어느 정도 도움을 줄 수 있는지를 분석하였다.

1. 서론

현재까지 추천시스템에 대한 연구가 활발하게 진행되고 있으며, 아마존(Amazon), 시디나우(CD Now) 등 해외의 우수한 사이트뿐만 아니라 삼성물, 한솔CS클럽 등 국내의 일부 쇼핑몰에서도 널리 적용되고 있다. 이런 추천시스템의 핵심은 추천알고리즘에 있으며, 협업필터링(Collaborative Filtering, CF)은 현재 가장 선호되고 있는 추천알고리즘으로 영화, 웹 사이트, 상품, 뉴스 등 여러분야에서 사용

되고 있다(Billsus & Pazzani, 1998; Cho et al., 2002; Cho & Kim, 2004; Hill et al., 1995; Lawrence et al., 2001; Resnick et al., 1994; Sarwar et al., 2000). 협업필터링은 목표고객과 유사한 기호를 가진 이웃고객의 기호와 매칭시켜 이웃이 구매한 상품 중 구매할 가능성이 가장 높은 상품을 추천한다. 그러나 협업필터링은 다른 알고리즘보다 훨씬 우수함에도 불구하고 다음과 같은 문제점이 존재한다(Cho et al., 2002; Cho & Kim, 2004; Sarwar et al., 2000). 첫째, 데이터의 희박성(sparsity)이다. 웹사이트에서 취급하는 컨텐

츠가 기하급수적으로 증가함으로써 고객의 선호도가 입력되지 않은 콘텐츠가 상대적으로 많아지고 있다. 따라서 유사한 이웃을 찾는 과정에서 아주 적은 선호도 정보가 사용되므로 유사한 이웃 고객에 대한 신뢰도와 추천의 질이 매우 낮아질 수 있다. 둘째 시스템의 확장성(scalability)이다. 인터넷 쇼핑몰의 추천시스템은 일반적으로 수백만 이상의 상품 및 고객을 처리한다. 하지만 고객과 상품의 수가 증가함에 따라 연산량이 기하급수적으로 늘어나기 때문에 실시간 추천을 목적으로 하는 상품추천시스템은 확장성문제에 직면하게 되었다(김재경 외, 2003; Sarwar et al., 2000).

이러한 희박성 문제와 확장성 문제를 해결하기 위하여 협업 필터링에 관한 많은 연구가 진행되어 왔다(김재경 외, 2004). 하지만 추천시스템은 이와 같은 데이터의 희박성과 시스템의 확장성 문제뿐만 아니라 추천절차의 블랙박스로 인한 문제가 존재할 수 있다. 추천절차의 블랙박스 문제는 추천시스템이 추천결과를 제공하지만 '왜 그런 결과가 도출되었는지?', '어떤 데이터를 이용 하였는지?', '어떻게 추천이 이루어 졌는지?'와 같은 추천과 관련된 정보를 제공해 주지 못하기 때문에 발생한다. 또한 협업필터링 기반의 추천시스템은 모든 고객의 현재 요구를 반영하지 못하고, 극단적으로 희박하거나 불완전한 데이터로 추천이 계산되기 때문에 추천이 정확할 수도 있지만, 경우에 따라서는 매우 잘못된 추천정보를 제공할 수 있다. 따라서 협업필터링 기반 상품추천시스템이 제공한 추천이 부정확한 추천이라면, 추천결과만을 믿고 상품을 선택한 고객은 많은 손실을 입게 된다. 이런 이유로 고객들은 추천을 완전히 신뢰하지 못하고 있으며, 자동화된 협업필터링은 엔터테인먼트 도메인에서는 어느 정도 성공적이지만 위험도가 높은 도메인, 예를 들면 펀드, 허니문여행 추

천과 같은 도메인에서는 좋은 성과를 보이지 못하고 있다(Dahlen et al., 1998). 특히 위험도가 높은 상품 도메인에서 상품 추천을 할 때 추천 결과와 함께 설명 기능을 추가한다면, 고객의 의사결정에 더욱 큰 도움을 줄 것으로 기대된다. 하지만 현재까지 설명기능이 추가된 추천방법에 관한 연구는 Herlocker et al.(2000) 외에는 아직 부족한 상황이다.

따라서 본 연구에서는 기존의 협업필터링 기반 추천시스템에 여러 형태의 설명기능을 추가한 WebCF-Exp추천시스템을 개발하고 설명기능의 추가가 가져오는 성과의 증진과 설명의 유형에 따른 효과성의 차이를 실험을 통하여 검증하였다. 본 연구의 실험을 위하여 가상의 인터넷 쇼핑몰을 구현하였으며, 인터넷 쇼핑몰을 이용한 고객들을 대상으로 온라인 설문을 통해 설명기능을 이용한 상품추천이 기존의 상품추천보다 사용자들의 구매 의사결정에 어느 정도 성과가 있는지, 그리고 각기 다른 20개의 설명인터페이스를 이용한 상품 추천이 고객들의 구매 의사결정에 영향을 미치는 정도를 분석하였다.

본 연구의 구성을 살펴보면 다음과 같다. 2장에서는 상품추천시스템과 설명기능에 관련된 기존의 연구들에 대해서 살펴보고, 3장에서는 본 연구에서 제안하고자 하는 WebCF-Exp추천시스템의 구조를 소개하고 각 구성 요소들에 대해 설명한다. 4장에서는 설명기능을 평가하기 위한 설문과정과 연구절차 그리고 실험 및 결과에 대해 설명하고, 5장에서는 연구의 결론을 제시한다.

II. 관련 연구

2.1 협업필터링 기반 상품추천시스템

협업필터링은 웹을 기반으로 하는 인터넷 쇼핑물에서 이용되고 있는 성공적인 상품추천기법중의 하나로써, 목표고객과 유사한 구매이력을 보이는 이웃 고객들의 상품에 대한 선호를 바탕으로 목표고객에게 유용한 상품을 추천하는 방법이다(김재경 외, 2003; Herlocker et al., 1999; Resnick et al., 1994; Sarwar et al., 2000). 일반적으로 이러한 협업필터링 기반 상품추천과정은 크게 입력 데이터 구성, 이웃 집단 탐색, 추천 상품 결정 단계로 나뉘 볼 수 있으며, 이러한 과정을 자세히 살펴보면 다음과 같다.

- 1) 입력 데이터 구성: 협업필터링 기반 상품추천 시스템에서의 입력데이터는 보통 n 개의 상품에 대하여 m 명 고객의 구매 트랜잭션 집합으로 구성되며, 보통 $n \times m$ 의 고객-상품 행렬 R 로 표현될 수 있다.
- 2) 이웃 집단 탐색: 고객 간의 유사도를 계산하여 이웃 집단을 탐색하는 과정이다. 두 고객 a 와 b 의 유사도를 측정하는 방법으로써 피어슨 상관계수(Pearson Correlation), 코사인(Cosine) 등을 사용한다.
- 3) 추천 상품 결정: 상품 추천을 위한 마지막 단계로서 설정된 이웃 집단으로부터 상위 N 개의 추천 상품 목록을 이끌어 내는 단계이다.

그러나 이러한 협업필터링 기반 추천시스템은 인터넷 쇼핑물의 상품과 고객수의 급속한 증가로 인해

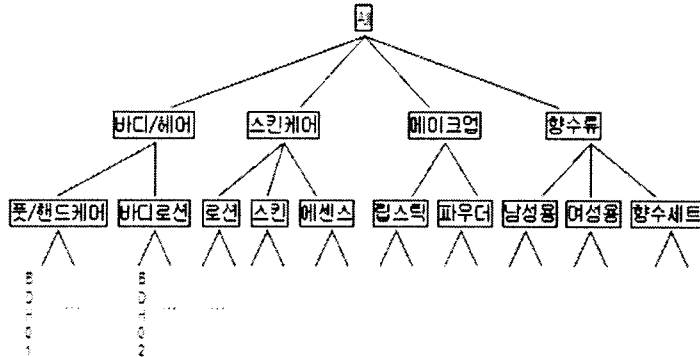
입력 데이터의 희박성과 시스템 확장성 문제를 노출시키고 있을 뿐만 아니라 추천절차의 블랙박스로 인한 문제가 존재할 수 있다(Billsus & Pazzani, 1998; Dahlen et al., 1998; Herlocker et al., 2000; Sarwar et al., 2000).

2.2 웹마이닝과 상품계층도를 이용한 추천시스템

기존의 협업필터링 기반 추천시스템의 문제점을 해결하기 위하여 Cho & Kim(2004)은 웹마이닝과 상품계층도를 이용한 추천시스템을 제안하였다. 웹마이닝(Web Mining)은 웹 사이트에서 고객이동 경로 즉 클릭스트림 정보를 가지고 있는 웹로그(Web log)로부터 고객들의 접속관계·패턴·규칙 등을 찾아내고 모형화해 유용한 마케팅 정보로 변환시키는 일련의 과정을 말한다(Mobasher et al., 2000). 웹마이닝의 전체 프로세스는 데이터 전처리와 패턴분석과정 두 가지로 분류된다. 데이터 전처리는 획득한 웹로그 데이터를 분석에 적합한 형태로 정제하여 불필요한 데이터를 제거함으로써 원천데이터의 용량을 감소시킬 수 있다(Cooley et al., 1999). 두번째, 패턴분석과정은 연관성규칙, 연속적인 패턴 발견 등을 이용하여 고객의 쇼핑행위에서 구매 패턴을 발견할 수 있다(Mobasher et al., 2000). 이러한 웹마이닝을 통하여 기존에 발견할 수 없었던 고객 선호도를 보다 많이 확보할 수 있으므로, 협업필터링 기반 상품추천시스템의 데이터 희박성 문제는 해결될 수 있다(Cho & Kim, 2004).

상품계층도는 개별상품을 상품클래스로 분류하고 이들 상품 클래스를 다시 추상 개념이 좀더 높은 상품 클래스로 분류한 계층 구조를 말하며 일반적으로 개별 상품, 상품 카테고리(Category), 상품군 등의 순서로 형성된다. <그림 1>은 상품계층도의 예를 보

〈그림 1〉 상품계층도



여주고 있다. 최근 많은 데이터 마이닝 관련 연구들이 데이터 분석에서의 상품 계층도의 필요성 및 중요성을 지적하고 있다(Han & Fu, 1999; Lawrence et al., 2001). 상품계층도상에서 유사한 선호도 패턴을 갖는 개별상품들을 특정 상품군으로 군집화하여 입력 데이터의 차원을 축소하면 이웃 집단 탐사 과정에서의 계산속도가 향상되어 시스템 확장성 문제 해결에 도움이 된다(김재경 외, 2004; Cho & Kim, 2004).

2.3 상품추천시스템의 설명기능

당신의 친한 친구가 새로운 상품을 추천하였다고 가정하자. 그 친구의 추천을 받아들일 때, 당신은 그의 과거 추천품질과 추천한 도메인에 관한 지식을 고려할 것이고, 그 추천에 의심이 간다면 추천에 관한 이유를 물을 수 있다. 이와 같은 추천에 관한 근거 및 이유 등의 관련 지식은, 추천을 받은 후 그 정보를 분석하고 고려하여 구매할 때의 의사결정에 도움이 될 것이다. 그러나 현재의 대부분 추천시스템은 추천결과만을 제공함으로써 몇 가지 오류가 발생할 수 있는데, 대표적인 오류로는 모델/프로세스 오

류(Model/Process Errors)와 데이터 오류(Data Errors) 두 가지로 분류할 수 있다(Herlocker, 2000).

먼저 모델/프로세스 오류는 추천시스템이 사용자의 요구사항을 고려하지 않고 추천모델이나 추천방법만으로 추천을 할 경우 발생한다. 예를 들면, 고객 프로파일상의 평가정보만을 이용하여 추천을 형성하는 추천시스템이 있다고 가정하고 어떤 고객 Kim이 여자친구와 크리스마스에 볼 영화를 선택하려고 영화 추천시스템을 이용한다고 하자. 이때, 일반적으로 추천시스템은 Kim의 과거의 평가정보만을 기반으로 과거에 선호했던 공포영화와 관련해서 추천을 형성한다. 따라서 추천시스템은 Kim의 현재 상황을 고려하지 않은 채 신작 공포영화만을 추천함으로써 Kim에게 불필요한 추천을 제공하는 추천오류를 발생하게 된다. 두번째, 데이터 오류는 부정확한 데이터를 이용하여 추천 할 때 발생할 수 있다. 이러한 경우 사용자는 이유도 알지 못한 채 잘못된 추천만으로 의사결정을 할 수 있으므로, 추천의 오류가 발생할 가능성이 높아진다.

이와 같은 추천의 오류를 해결하기 위해 Herlocker (2000)는 상품추천에서 발생하는 오류를 조절하고, 고객의 의사결정에 도움을 주기 위한 메커니즘으로

설명(Explanation)기능을 제안하였다. 설명기능은 추천이전의 추천절차와 데이터를 설명함으로써 추천 과정의 투명성을 제공한다. 따라서, 추천시스템에 설명기능을 추가함으로써, 당위성(Justification), 사용자 참여(User Involvement), 교육기능(Education), 수락(Acceptance)등의 이익을 얻을 수 있다고 하였다. 기존의 전문가시스템(Expert System)에서도 이러한 설명기능을 가지고 있었는데, 예를 들면, 사용자에게 시스템이 "어떻게 결론을 내렸는가?", 또는 "왜 특별한 데이터를 필요로 하는가?"와 같이 결론이나 중간결과에 도달하게 된 추론과정을 사용자에게 설명해주는 기능을 가지고 있었다(Kim & Kim, 1991). 즉 추론의 수행과정에 대해서 사용자에게 설명해줌으로써 사용자의 의문을 해소하고 사용자의 신뢰를 얻을 수 있게 해 준다. 이와 같이 설명기능은 질의응답 수행시 현재 진행하고 있는 질문에 대해 사용자가 이해하지 못하거나 설명을 필요로 할 경우, 사용자의 이해를 돕기 위해 자세한 설명을 해주는 모듈이라 할 수 있다.

Herlocker(2000)에 의하면, 설명기능에 이용되고 있는 모델은 크게 화이트박스 모델과 블랙박스 모델로 구분된다. 화이트박스 모델은 추천절차와 데이터를 설명해 줌으로써 고객이 의사결정을 내리는데 도움을 주는 방식이다. 자동화된 협업필터링 추천시스템은 '고객 프로파일 입력', '유사한 프로파일로 이웃 형성', '이웃들의 정보를 조합하여 추천을 형성'과 같은 3단계에 의해 추천을 제공한다. 추천시스템을 구현할 때 3단계 과정은 좀더 세부적인 단계로 나누어 지지만 일반적으로 사용자들은 인식하지 못한다. 만약 고객들이 3단계의 추천수행과정을 이해한다면 추천시스템의 전반적인 과정을 인지할 수 있을 것이다. 블랙박스 모델은 추천 메커니즘과 독립적으로 추천시스템의 입력 값과 출력 값, 상품추천에 영향

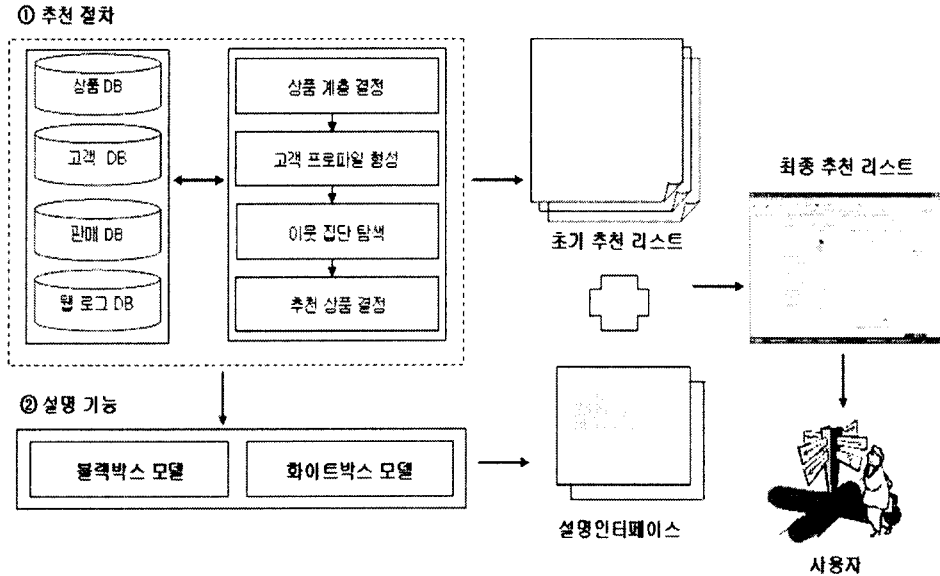
을 줄 수 있는 요소들에만 초점을 맞춘다. 블랙박스 모델의 예로는 '높은 정확성이 검증된 추천시스템은 이 상품이 고객님의게 적합하다고 판단 하였습니다'와 같은 시스템의 과거 성능, 고객 만족도의 표시, 전문가의 논평 또는 '이 상품은 특가기획상품으로 행사기간 중 40%가 할인 됩니다'와 같은 마케팅 정보 등이 있다.

본 연구에서 구현한 설명 인터페이스는 이 두 가지 모델을 다 활용하고 있다. 여러 설명 인터페이스 중 성공적인 설명 인터페이스는 고객이 추천된 제품을 구매하는데 도움을 줄 수 있으며, 추천 받은 고객이 추천의 장점과 단점을 판단하는데 도움을 줄 수 있다. 또한 고객이 설명기능의 유용성을 인식하여 추천 시스템을 지속적으로 사용하는데도 도움을 줄 수 있을 것이다. 따라서 본 연구에서는 다양한 설명 인터페이스를 구현한 후 온라인 설문문을 통해 이와 같은 사실을 검증하고자 한다.

III. 설명기능이 추가된 상품추천방법

WebCF-Exp추천방법은 <그림 2>와 같이 추천절차와 설명기능으로 구성되어 있다. 추천절차는 상품 계층결정, 고객 프로파일 형성, 이웃 집단 탐색, 추천 상품 결정의 4단계로 구성되며, 추천절차를 통해 고객에게 제공할 상품추천 리스트를 형성한다. 설명기능은 블랙박스 모델과 화이트박스 모델로 구성되어 있고, 설명인터페이스를 형성한다. 추천시스템은 최종적으로 상품추천리스트와 설명 인터페이스를 조합하여 최종 추천리스트를 사용자에게 제공한다.

〈그림 2〉 WebCF-Exp 추천방법



3.1 WebCF-Exp 추천절차

WebCF-Exp추천시스템의 추천절차는 Cho & Kim(2004)이 제안한 방법을 이용하고 있다. 즉 첫 단계인 '상품계층결정'단계는 데이터베이스 내의 모든 상품을 마케팅 전문가가 특정그룹으로 분류·재구성하여 입력 데이터를 감소시킨다. '고객 프로파일 형성'단계는 인터넷 쇼핑물에서 개별고객의 쇼핑행위를 추적하여 얻은 데이터를 통해 목표고객의 상품 선호도 정보를 발견하고, 분석하여 고객 프로파일을 형성한다. '이웃 집단 탐색'단계는 형성된 고객 프로파일을 이용하여 고객 간의 유사도를 계산하고 이를 기반으로 목표고객과 유사한 성향을 가진 고객을 선택하여 이웃을 형성한다. 마지막으로 '추천상품결정' 단계에서는 형성된 이웃들의 쇼핑 행위를 기반으로 선호도가 높은 상위 N개의 상품을 선택하여 추천목록을 형성한다.

3.2 WebCF-Exp 설명기능

기존 협업필터링 기반 상품추천시스템과 달리 추천 절차를 실행한 후 추천에 적합한 설명인터페이스를 형성하는 단계이다. 설명인터페이스는 설명기능을 고객에게 제공할 화면을 구현한 것으로 고객이 추천결과에 대한 배경을 알고자 할 때 제공함으로써 고객의 의사결정에 도움을 줄 것이다. 따라서 본 연구에서는 화이트박스 모델과 블랙박스 모델을 참조하여 <표 1>과 같이 20개의 설명인터페이스를 구현하였다.

이러한 설명기능을 제공하는 20개의 설명인터페이스는 부록에 나타나 있다. 본 연구에서는 이러한 20개의 설명기능을 화면으로 구현한 후 온라인설문을 통해 설명기능의 추가가 가져오는 성과의 증진과 20개의 설명인터페이스 유형에 따른 효과성의 차이를 실험을 통하여 검증하였다.

〈표 1〉 설명인터페이스 유형

	설명인터페이스 유형
1	상품조회 비율(그래프)
2	상품조회 비율(기술)
3	상품조회 비율(테이블)
4	장바구니 담은 비율(그래프)
5	장바구니 담은 비율(기술)
6	장바구니 담은 비율(테이블)
7	구매 비율(그래프)
8	구매 비율(기술)
9	구매 비율(테이블)
10	유명 매체의 추천(기술)
11	추천시스템의 예측(기술)
12	구매한 상품과의 유사 정도(기술)
13	이웃의 수(기술)
14	이웃의 구매비율(기술)
15	목표고객 상품군별 선호도(그래프)
16	목표고객 상품군별 선호도(기술)
17	목표고객 상품군별 선호도(테이블)
18	이웃의 선호도(5개의 별)
19	마케팅 정보1(기술)
20	마케팅 정보2(기술)

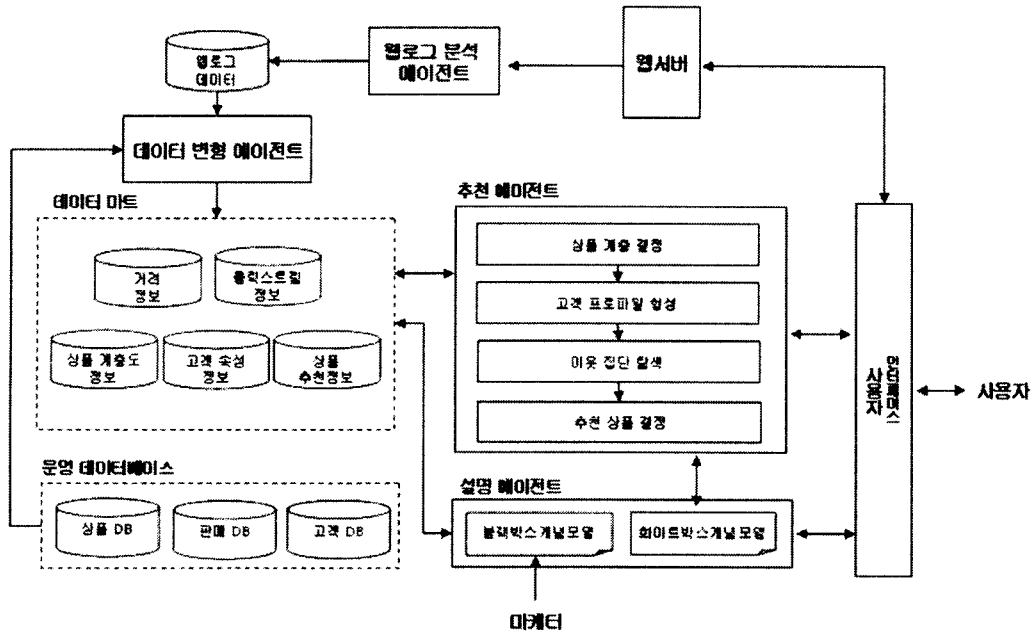
IV. WebCF-Exp 추천시스템

WebCF-Exp시스템은 본 연구자(Cho & Kim, 2004))들이 이미 개발한 WebCF-PT시스템과 동일한 추천방법을 사용하고 있으며, 설명기능을 추가하여 개발한 추천시스템이므로 추천 방법에 대한 자세한 설명은 기존 연구를 참조하기 바란다. WebCF-Exp추천시스템의 전체 구조는 〈그림 3〉과 같으며, 웹로그 분석 에이전트(Web Log Analysis Agent),

데이터 변형 에이전트(Data Transformation Agent), 추천 에이전트(Recommender Agent), 그리고 설명 에이전트(Explanation Agent)로 구성되어 있다.

입력 데이터로는 운영 데이터베이스에 저장된 고객 데이터, 상품 데이터, 판매 데이터와 웹로그 데이터이며, 추천시스템의 최종 결과로 개인화된 상품추천리스트와 추천을 설명하는 설명인터페이스를 제공한다. 각 구성요소들의 기능을 살펴보면, 먼저 웹로그 분석 에이전트가 방대한 양의 웹로그 파일을 전처리하여 웹로그 데이터에 저장하면, 데이터 변형

〈그림 3〉 WebCF-Exp추천시스템 구조도



에이전트는 운영 데이터베이스와 웹로그 데이터로부터 상품추천에 사용될 데이터를 추출, 정제, 변형하여 데이터 마트에 저장한다. 추천 에이전트는 추천 알고리즘을 통해 고객별 개인화된 상품추천 리스트를 작성하고 설명기능 구현에 필요한 고객 선호도, 추천에 사용된 데이터, 그리고 실시간 추천을 제공하기 위한 추천리스트를 데이터 마트에 저장한다. 설명 에이전트는 추천에 관한 이해를 돕기 위한 설명인터페이스를 작성한다. 작성된 추천리스트와 설명인터페이스는 사용자인터페이스를 통해 고객에게 제공된다.

4.1 시스템 구성 에이전트

- **웹로그 분석 에이전트:** 웹 서버에 있는 텍스트 형태의 웹로그 파일은 곧바로 유용한 정보가 되지

못하므로 전처리를 하며, 그 결과 산출된 데이터를 데이터베이스에 저장하게 된다. 데이터를 적재하기 전의 전처리 과정은 데이터 정제(Data Cleaning), 사용자 확인(User Identification), 세션 식별(Session Identification), 클릭스트림 경로완성, 포매팅(Formatting), 데이터 로딩(Data Loading)으로 구성된다.

- **데이터 변형 에이전트:** WebCF-Exp추천시스템의 데이터 마트를 유지 관리하기 위해 운영 데이터베이스와 웹로그 데이터베이스로부터 자료를 추출, 정제하고 추천시스템에 적용하기 적합한 형태로 데이터를 변환시키는 과정을 수행한다. 데이터 변형 에이전트는 운영 데이터베이스의 고객 정보로부터 분석에 필요한 데이터를 추출하여 데이터 마트의 고객 속성 정보에 저장하고 운영 데이터베이스의 판매 데이터와 상품 데이터로부터 상

품 정보, 상품 클래스 정보, 구매 빈도를 추출하여 거래 정보에 저장한다. 그리고 웹로그 데이터베이스로부터 고객별 쇼핑 단계 정보와 상품코드를 추출하여 클릭스트림 데이터에 저장한다. 이때 상품코드가 없는 로그정보는 삭제하고 웹 사이트의 URL를 쇼핑 행위의 각 단계로 구분하여 각 단계마다 특정 상품이 발생한 빈도수를 합산한 후 저장한다. 상품계층도 정보는 마케팅 담당자나 도메인 관리자에 의해 미리 정의된 상품카테고리를 상품계층에 따라 분류하여 저장한다.

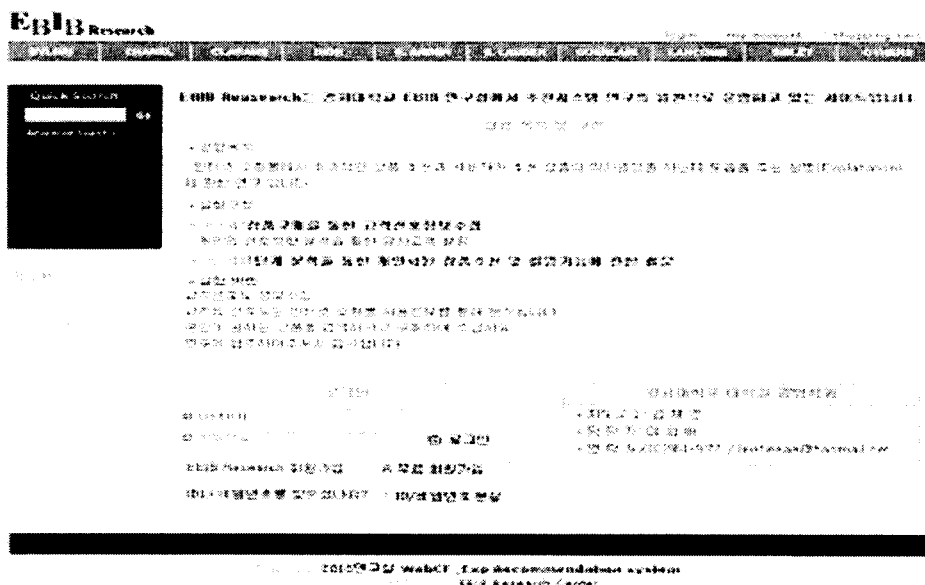
- **추천 에이전트:** 추천 에이전트는 데이터 마트의 상품 관련 정보와 고객 관련 정보를 이용하여 구현된 개인화된 추천목록을 생성한다. 그리고 상품 추천과정에서 발생한 고객 프로파일 정보, 이웃고객 정보, 고객 선호도 정보, 추천상품 정보를 데이터 마트에 저장한다.
- **설명 에이전트:** 설명 에이전트는 화이트박스 모

델과 블랙박스 모델을 기반으로 데이터 마트에 저장되어 있는 선호도 정보, 추천 상품 정보, 이웃 정보와 마케터의 입력 값 그리고 추천절차를 이용하여 설명인터페이스를 구현한다.

4.2 설명기능이 추가된 상품추천의 예

이상에서 소개한 WebCF-Exp추천시스템이 과연 실제로 적용했을 때 어떤 결과를 보일 것이며, 또 설명기능을 추가 했을 때 추천상품에 대한 사용자의 의사결정에 도움을 줄 수 있는지 확인해보기 위해, 본 연구에서는 WebCF-Exp추천시스템을 프로토타입 형태로 개발하고 온라인 가상 쇼핑물 EBIB Research 쇼핑물을 <그림 4>와 같이 구현하였다. 이 사이트는 고객이 고위험을 내제한 상품을 추천 받았을 때 설명기능이 구매의사결정에 미치는 영향을 알아보기 위하여 상품도메인을 고가의 명품으로

<그림 4> EBIB Research 쇼핑물 초기화면



선정하였다.

〈그림 5〉는 고객 Kim에게 제공된 개인화된 추천 상품 리스트의 예를 보여주고 있다. 추천에이전트는 이렇게 계산된 협업필터링 기반 정보인 이웃정보, 추천상품 정보를 데이터 마트에 저장한다. 이 정보를 기반으로 설명에이전트는 설명인터페이스를 구현한다.

Kim은 추천된 상품 중 2번 '석세스 아이크림'을 자세히 보기 위하여 〈그림 5〉의 상품 이미지를 클릭

하면, 〈그림 6〉과 같은 상품상세보기 화면으로 이동한다. 기존 상품추천시스템의 상품상세보기 화면과 다르게 설명기능을 제공하기 위하여 [이 상품을 추천한 이유]버튼이 추가 되었다. Kim은 자신이 선호하는 상품이지만 고가이고, 왜 이 상품이 추천되었는지 이유를 알고 싶어할 것이다. 그때, 새로 추가된 설명기능 버튼을 발견하고, 버튼을 누르면 〈그림 7〉과 같은 설명인터페이스 팝업창이 뜨고 Kim에게 추천한 이유를 설명한다.

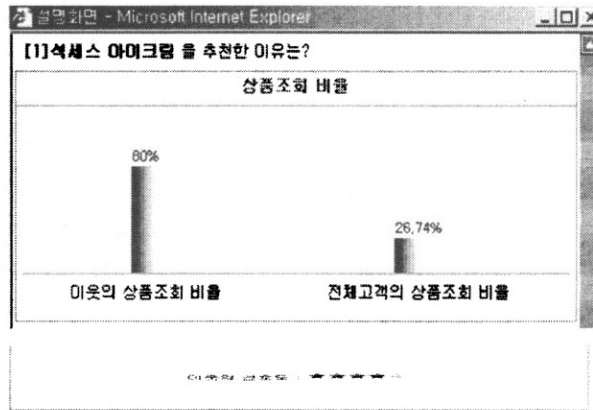
〈그림 5〉 고객 Kim에게 제공된 상품추천리스트

1		상품명 알마니 시계 AR-0224 가격 219,000 제조사 알마니
2		상품명 석세스 아이크림 가격 112,000 제조사 결랑
3		상품명 Ferragamo Shose 가격 615,000 제조사 페라가모
4		상품명 프레췌지용 푸리파임 토니 [200ml/6.7oz] 가격 34,400 제조사 샤넬
5		상품명 프라다56-741 가격 890,000 제조사 프라다
6		상품명 크리스찬 디올 프레스타지 20+1 뉴트리 - 리스토링 크림 (50ml) 가격 288,000 제조사 크리스찬디올
7		상품명 오흐리 PRO M프레임(코팅) 가격 250,000 제조사 오흐리
8		상품명 콘스텔레이션 올바 남성용(골드-1212, 10) 가격 2,450,000 제조사 오메가
9		상품명 Absolute Night Recovery Treatment 가격 137,500 제조사 알코
10		상품명 협희 XR60/60 (엑스트라 바이탈 리스토링 에센스) 30ml 가격 70,200 제조사 크리스찬디올

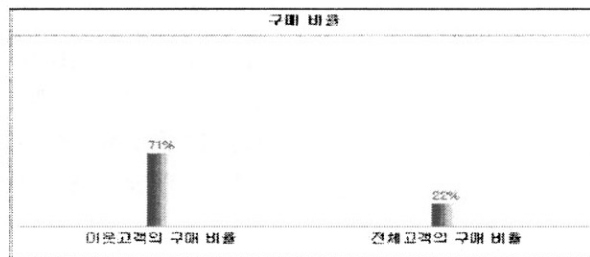
〈그림 6〉 고객 Kim에게 추천된 상품의 상품상세보기 화면



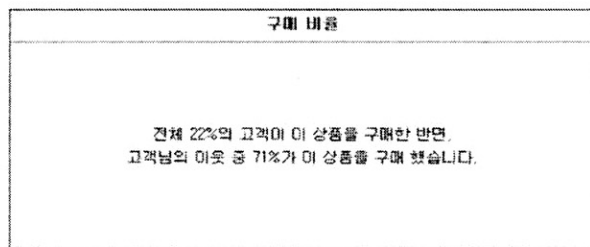
〈그림 7〉 설명인터페이스



〈그림 8〉 설명인터페이스 예



(a) 구매 비율(그래프)



(b) 구매 비율(기술)

구매 비율	
이웃고객의 구매 비율	71%
전체고객의 구매 비율	22%

(c) 구매 비율(테이블)

〈그림 7〉의 설명인터페이스는 이웃의 참조빈도 즉 상품조회 비율을 보여주고 있다. 전체고객 중 26.7%가 이 상품을 상세보기한 반면, Kim과 유사한 선호도를 가진 이웃고객의 80%가 이 상품을 상세보기하였고 이웃의 선호도는 5점 만점에 4점이다. 따라서 Kim과 유사한 선호도를 가진 고객들은 이 상품에 많은 관심을 가지고 있다는 것을 알 수 있으며, 이 설명인터페이스는 이웃고객의 선호도가 높게 나타났으므로 Kim에게 이 상품을 추천한 이유에 대한 설명을 제공한다고 볼 수 있다.

〈그림 8〉은 본 연구에서 구현한 20개의 설명인터페이스 중 대표적 표현인 그래프, 기술(description), 테이블 형태의 예를 보여주고 있다.

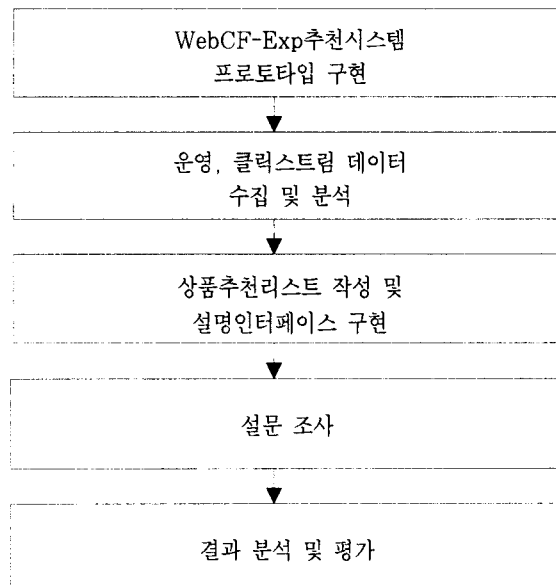
V. 실험 및 평가

지금까지 기술한 추천시스템 설명기능의 성능 평가 및 설명유형의 차이를 검증하기 위해서 WebCF-Exp추천시스템, EBIB Research 온라인 쇼핑몰, 그리고 20개의 설명인터페이스를 구현하였으며, 온라인 설문을 실시하였다. 본 연구의 실험 절차는 〈그림 9〉와 같은 절차에 의해 수행되었다.

5.1 실험절차 및 방법

실험을 위해 168명의 피실험자들에게 사이버 머니를 나누어주고 EBIB Research 쇼핑몰을 2003년 10월 2일부터 11월 26일까지 이용하게 하였다. 피실험자들의 프로파일이나 로그파일을 분석한 선호도를 바탕으로 상품을 추천하고, 11월 18일부터 11

〈그림 9〉 실험 절차



월 26일까지 피실험자들의 구매데이터를 분석하여 WebCF-Exp추천시스템의 성능을 평가하였다. 또한 이 기간 동안 설명인터페이스를 통하여 추천을 받은 고객들을 대상으로 설명기능에 관한 설문조사를 실시하였다.

5.1.1 WebCF-Exp추천시스템 프로토타입 구현

WebCF-Exp추천시스템의 데이터베이스 시스템으로 MS-Access를 사용하였고, 각 에이전트와 구성요소들은 Visual Basic6.0을 사용하여 개발하였다. 그리고 가상 인터넷 쇼핑몰인 EBIB Research 쇼핑몰은 데이터베이스시스템으로 MS-SQL Server 2000을 사용했고, WindowsNT2000 기반의 IIS5.0 웹 서버와 ASP(Active Server Page)를 이용하여 개발하였다.

5.1.2 데이터 수집 및 분석

- 웹로그 데이터: MS IIS가 제공하는 여러 로그파일 중 W3C Extended 로그파일을 사용하였고, 웹 서버로부터 68개의 로그파일을 분석하였다. 웹로그 분석 에이전트를 이용하여 <time, user_id, product-id, shopping_step>의 형태로 데이터를 추출하고, 웹로그 데이터에 입력한다. 그리고 웹로그 변형 에이전트를 이용하여 데이터를 변형한 후 데이터 마트에 저장하였다.
- 상품 데이터: EBIB Research 쇼핑몰은 287개의 상품으로 거래가 이루어진다. 상품계층은 루트노드를 제외한 3개의 계층으로 구성되어 있다. 상위계층은 7개의 상품클래스가 있고 다음계층은 37개의 상품 클래스로 구성되어 있고 맨 아래 계층은 287개의 개별상품으로 구성되어 있다.

- 추천 데이터: 고객에게 적시에 상품추천을 제공할 수 있도록 목표고객의 이웃인 30명의 고객ID를 저장하였다. 그리고, 개인별 상품추천 목록을 제공받을 고객 ID와 상품코드를 데이터 마트에 저장하였다.

5.1.3 상품추천 리스트 작성 및 설명인터페이스 구현

분석된 데이터를 이용하여 고객에게 개인화된 상품추천 목록을 제공한다. 그리고 설명모델을 기반으로 WebCF-Exp추천시스템의 추천절차와 이용된 데이터를 사용하여 20개의 설명인터페이스를 구현하였다.

5.1.4 설문조사

성공적인 설명인터페이스는 고객이 추천된 제품을 구매하도록 확신시키는 것으로써 고객이 설명기능의 유용성과 흥미를 느끼고 추천시스템을 지속적으로 사용할 것이라는 인식을 심어주는 것이다. 따라서 설명기능은 고객이 이해하기 쉽도록 구현되어야 하고 추천시스템에 설명기능이 추가되었을 때, 그것이 불필요한 정보가 아니라 고객의 추천의사결정에 도움이 되어야 한다. 따라서 본 연구에서는 WebCF-Exp추천시스템의 설명기능의 효과성과 20개의 설명 인터페이스 중에서 적합한 인터페이스를 알아보기 위하여 EBIB Research 쇼핑몰을 이용하여 온라인 설문조사를 실시하였다. 개인화된 상품추천을 받은 후 설명 인터페이스를 본 고객을 대상으로 2가지 질문을 하였다. 먼저 인터페이스의 이해도와 적합성을 알아보기 위하여 질문1을 물어보았다.

질문1. 설명인터페이스는 이해하기 쉬웠습니까?

앞서 구현한 20개의 설명인터페이스 중 무작위로 선정하여 상품과 관련된 설명을 제공하였고, 설명 인터페이스 이해 정도를 리커트(Likert) 5점 척도로 매우 어렵다(1), 어렵다(4), 보통이다(3), 쉽다(4), 매우 이해하기 쉽다(5)를 사용자가 선택하여 입력하게 하였다.

두번째로는 추가된 설명인터페이스를 본 후 구매 의사결정에 도움을 주었는지 알아보기 위해 다음과 같은 질문을 하였다.

질문2. 제시된 설명인터페이스가 추천상품 구매 여부에 어느 정도 도움이 되었습니까?

질문 1과 마찬가지로 5점 척도를 사용하여 본 질문에 대해 전혀 도움이 되지 않았다(1), 도움이 되지 않았다(2), 보통이다(3), 도움이 되었다(4), 많이 도움이 되었다(5)로 사용자의 의견을 물어보았다.

5.2 실험 결과

5.2.1 WebCF-Exp추천시스템의 성과

〈표 2〉는 WebCF-Exp추천방법과 기존 WebCF-PT추천방법의 성능을 비교한 결과이다. 4장에서도 언급하였듯이 WebCF-Exp추천방법은 기존 WebCF-PT추천방법과 동일한 추천환경으로 설명기능만 추가한 추천방법이다. 두 추천방법의 성능을 평가하기

위해 일반적으로 많이 사용되는 정확도(Precision)와 예측률(Recall)을 평가척도로 사용하였다(Sarwar et al., 2000). 정확도는 추천목록 중 몇 개의 상품을 고객이 구매하였는지를 나타내는 평가방법이고, 예측률은 고객이 구매한 상품 중에서 얼마나 많은 상품이 추천되었는지 평가하는 방법이다. 일반적으로 추천방법의 정확도와 예측률을 측정하기 위해서는 이웃의 수와 추천상품의 수, 추천상품 결정 방법을 정해야 된다. 여러가지 방법 중에서 어느 방법을 사용하는가는 분야별로 결과가 다르기 때문에 주로 휴리스틱한 방법으로 실험을 통하여 결정한다. 이 연구에서는 본 연구자(Cho & Kim, 2004)들이 실험을 통해 제안한 방법을 기반으로 이웃의 수는 30명으로, 추천상품의 수는 10개로, 그리고 추천상품 결정방법은 MFR(Most Frequently Referred product)추천방법으로 선택하였다. 자세한 내용은 참고문헌을 참고하기 바란다.

실험결과를 보면 본 연구에서 제안한 설명기능이 추가된 WebCF-Exp 추천시스템이 정확도와 예측률 모두 기존 WebCF-PT 추천방법 보다 높은 평균값을 얻었다. 이것을 통계적으로 분석하기 위해 쌍체 표본 평균검정(Paired Samples T-Test)을 실시하였다. 〈표 3〉은 쌍체 표본 통계량을 제시하고 있다. WebCF-Exp추천방법과 WebCF-PT추천방법의 표본수는 각각 112명 이고, 정확도와 예측률의 평균은 〈표 2〉와 동일하게 각각 0.196과 0.123, 0.495와 0.425로 WebCF-Exp추천방법의 정확도

〈표 2〉 추천방법의 성능 비교

	WebCF-Exp 추천방법	WebCF-PT 추천방법
정확도(Precision)	0.196	0.123
예측률(Recall)	0.495	0.425

〈표 3〉 쌍체 표본 통계량

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	WebCF-Exp_Precision	.19554	112	.20419	1.9294E-02
	WebCF-PT_Precision	.12321	112	.10221	9.6581E-03
Pair 2	WebCF-Exp_Recall	.49524	112	.41359	3.9081E-02
	WebCF-PT_Recall	.42525	112	.31951	3.0191E-02

와 예측률이 높은 것으로 나타났다.

〈표 4〉는 쌍체 표본 상관관계를 제시하고 있으며, WebCF-Exp추천방법과 WebCF-PT추천방법의 정확도와 예측률의 상관관계는 각각 0.700, 0.847이고 P값이 각각 0.000으로 유의수준 5%내에서 유의하고 상관관계가 높다고 할 수 있다.

〈표 5〉는 실제로 쌍의 차이(Paired Differences)가 나타나는지를 제시하고 있다. 결과 값을 보면,

WebCF-Exp추천방법과 WebCF-PT추천방법의 정확도가 차이가 있는지에 대한 검정값(Test Value)인 t값은 5.055이고, 자유도(df)는 111이며, P값은 0.000으로 유의수준 5%내에서 유의하다고 할 수 있다. 따라서 WebCF-Exp추천방법과 WebCF-PT추천방법의 정확도 평균 차이는 유의한 것으로 나타났다. 또한 WebCF-Exp추천방법과 WebCF-PT추천방법의 예측률 차이가 있는지에 대한 t값은 3.335

〈표 4〉 쌍체 표본 상관관계

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	WebCF-Exp_Precision & WebCF-PT_Precision	112	.700	.000
Pair 2	WebCF-Exp_Recall & WebCF-PT_Recall	112	.847	.000

〈표 5〉 쌍체 표본 검정

Paired Samples Test							
Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Pair 1	WebCF-Exp_Precision & WebCF-PT_Precision	7.2321E-02	.15140	1.4306E-02	5.055	111	.000
Pair 2	WebCF-Exp_Recall & WebCF-PT_Recall	6.9991E-02	.22211	2.0988E-02	3.335	111	.001

이고, 자유도는 111이며, P값은 0.001으로 유의수준 5%내에서 유의하다고 할 수 있으며, WebCF-Exp추천방법과 WebCF-PT추천방법의 예측률 평균 차이는 역시 유의한 것으로 나타났다. 따라서 설명기능이 추가된 WebCF-Exp추천방법이 더 우수한 성능임을 알 수 있다.

5.2.2 설명인터페이스의 효과성

20개의 설명인터페이스의 효과성을 알아보기 위해 온라인 설문을 통해 수집한 자료를 이용하여 빈도분석과 분산분석을 실시하였다. <표 6>은 온라인 설문조사에 참여한 사용자들의 인구 통계학적 빈도분포를 나타낸다.

우선, 성별 분포를 살펴보면, 여성이 55.95%, 남

성이 44.05%로 여성이 다소 많았고, 연령별로는 21~25세가 52.98%로 가장 많았다. 지역분포는 서울이 67.26%로 가장 많았고, 경기도, 경상도, 전라도, 충청도 순으로 나타났다. 인터넷 사용기간은 3년 이상이 85.71%로 대부분의 응답자가 인터넷 사용에 익숙하다는 것을 알 수 있었다. 응답자에 대한 상품추천관련 질문에 관한 분석결과는 <표 7>과 같다. 인터넷 쇼핑물로부터 상품추천을 받은 횟수를 보면 4회 이상이 49.40%로 높은 응답률을 보였으며 이와 대조적으로 상품추천을 받은 경험이 없는 응답자가 22.02%로 두 번째로 높은 응답률을 보였다. 그리고 추천 받은 상품의 구매 횟수는 1~2회 38.69%, 없음 29.17%, 4회 이상 19.64%, 3~4회 이상 12.50% 순으로 나타났다. 상품을 추천 받은 상품카테고리에 대한 응답은 복수 응답이므로 빈

<표 6> 응답자에 대한 빈도분포

	구 분	빈도	비율
성 별	여	94	55.95
	남	74	44.05
연 령	15~20세	51	30.36
	21~25세	89	52.98
	26~30세	26	15.48
	31세 이상	2	1.19
지역분포	경기도	44	26.19
	경상도	5	2.98
	전라도	3	1.79
	서울	113	67.26
	충청도	3	1.79
인터넷 사용기간	1년 미만	1	0.60
	1년 이상~2년 미만	9	5.36
	2년 이상~3년 미만	14	8.33
	3년 이상	144	85.71

〈표 7〉 응답자에 대한 상품추천관련 빈도분포

	구 분	빈도	비 율
상품추천 받은 횟수	없음	37	22.02
	1~2회	29	17.26
	3~4회	19	11.31
	4회 이상	83	49.40
추천 받은 상품의 구매 회수	없음	49	29.17
	1~2회	65	38.69
	3~4회	21	12.50
	4회 이상	33	19.64
추천 받은 상품카테고리	서적/잡지	68	17.44
	의류/패션제품	56	14.36
	화장품/감화류	51	13.08
	문화상품(영화/콘서트등)	41	10.51
	가전제품	38	9.74
	컴퓨터 관련제품	36	9.23
	생활용품	22	5.64
	없음	22	5.64
	스포츠/레저용품	14	3.59
	기타	14	3.59
	여행/서비스	8	2.05
	식품/건강	7	1.79
	은행/재무서비스	7	1.79
	꽃	4	1.03
	농수산물	2	0.51

도와 백분율이 다른 측정치와 다르다. 추천 받은 상품카테고리를 보면 서적/잡지가 17.44%로 가장 높은 순위를 차지했고, 의류/패션제품 14.36%, 화장품/감화류 13.08%, 문화상품 10.51% 등의 순으로 나타났다.

질문1 (설명인터페이스는 이해하기 쉬웠습니까?)

에 대한 평가점수 평균과 표준편차는 다음 〈표 8〉과 같다. 3위까지의 순위를 보면 이웃의 선호도(5개의 별), 마케팅 정보(기술), 구매비율 (그래프)이 추천된 상품의 추천이유를 이해하는데 높은 점수를 얻었다. 그리고, 상위 10위까지의 설명을 보면 마찬가지로 이웃 고객의 과거 구매와 관련된 정보와 마케팅

정보로 구성되어 있다. 즉 이것은 자신과 비슷한 선호도를 가진 고객이 어떻게 그 제품에 대해 평가했고 어떤 관심을 가지고 있는지에 관한 정보가 의사결정을 할 때 중요한 요인으로 작용하며 고객들은 그 정보를 쉽게 이해한다는 것을 알 수 있다. 또한 고객들의 구매에 직접적으로 도움이 되는 마케팅 정보도 고객이 추천을 설명하는데 영향을 준다는 것을 알 수 있다.

높은 점수를 얻은 정보를 보면 그래프나 기술로 표현되어 있다. 즉, 고객들이 설명을 쉽게 이해하는

데는 요약된 정보를 해석, 이해해야 하는 테이블보다는 한눈에 볼 수 있는 그래프나 내용을 쉽게 풀어서 설명한 기술이 고객의 이해를 돕는데 더 유용한 것으로 나타났다. 예를 들어 부록의 설명7, 설명8, 설명9는 구매 비율을 설명한 설명인터페이스이다. 그래프로 설명인터페이스를 작성한 설명7은 3번째로 높은 점수를 얻은 반면, 똑 같은 정보를 기술로 표현한 설명8은 17위를 차지하였고 테이블형태의 설명9는 18위를 차지하였다. 똑 같은 정보가 표현 형식에 따라 서로 상이한 응답 결과를 보인 것을 알

〈표 8〉 설명인터페이스의 이해정도

	설명인터페이스	평균	표준편차
1	이웃의 선호도(5개의 별)	3.806	0.992
2	이 제품은 특가기획상품으로 행사기간 중에 정가의 40%가 할인 됩니다.(기술)	3.609	1.060
3	구매 비율(그래프)	3.536	1.069
4	장바구니담은 비율(그래프)	3.472	1.035
5	이 상품을 추천 받은 고객 중 64%가 구매 하였습니다.(기술)	3.443	1.127
6	이 제품은 과거 고객님이 구매한 상품들과 유사 합니다.(기술)	3.417	1.229
7	이 제품은 지난 주에 가장 많이 구매된 상품으로 이주의 상품으로 선정 되었습니다.(기술)	3.277	1.005
8	상품조회 비율(그래프)	3.261	0.993
9	장바구니담은 비율(기술)	3.247	0.950
10	고객님의 상품 분류별 선호도(그래프)	3.217	1.260
11	장바구니담은 비율(테이블)	3.210	1.028
12	상품조회 비율(테이블)	3.207	0.942
13	이 제품은 고객님과 유사한 선호도를 가진 30명의 구매정보를 기반으로 추천 되었습니다.(기술)	3.173	1.104
14	이 제품은 주요 일간지와 유명잡지에 여러번 소개되었습니다.(기술)	3.155	1.253
15	상품조회 비율(기술)	3.152	0.969
16	고객님의 상품 분류별 선호도(테이블)	3.149	1.051
17	고객님의 상품 분류별 선호도(기술)	3.138	1.197
18	구매 비율(기술)	3.100	1.150
19	구매 비율(테이블)	3.049	1.206
20	높은 정확성이 검증된 추천 시스템은 이 상품이 고객님이 적합할 것이라고 예측 했습니다.(기술)	3.023	1.116

〈표 9〉 설명인터페이스 유형에 따른 이해정도 분석결과

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	88.048	19	4.634	3.913	.000
Within Groups	2186.428	1846	1.184		
Total	2274.476	1865			

수 있었다. 기술로 설명한 설명인터페이스 중 전문 용어보다는 평상시 자주 듣던 문구의 이해도가 높았으며, 20번과 같이 명확하지 않고 모호한 표현은 고객들의 이해를 돕는데 도움이 되지 않는다는 것을 알 수 있다.

이것을 통계적으로 분석하면, 〈표 9〉와 같다. 〈표 9〉는 어떤 유형의 설명인터페이스가 효과적인가를 통계적으로 검증하기 위해 설명인터페이스 이해도에 대해 분산분석(ANOVA)을 실시한 결과를 보여 준다. 분석결과를 해석하면, F값은 3.913이고 P값이 0.000으로 유의수준 5%내에서 매우 유의하다고 할 수 있다. 그러므로 설명인터페이스 유형에 따라 사용자 이해정도가 달라진다고 할 수 있다.

추가로 사후분석을 한 결과, 설명유형 18번과 설명유형 2, 3, 6, 8, 9, 10, 11, 13, 16, 17번과는 통계적으로 유의한 차이가 있고, 설명유형 20번과 설명유형 2, 11번도 통계적으로 유의한 차이가 있는 것으로 나타났지만, 나머지 설명유형들 간의 차이는 통계적으로 없는 것으로 나타났다. 이러한 결과에서 알 수 있는 사실은 피실험자들에게 이해하기 쉬운 설명유형으로 이웃의 선호도(별표시)와 전체 고객들이 가장 많이 구매한 특정상품에 대한 마케팅 정보로 나타났다. 하지만 다른 형태의 설명유형은 피실험자들에게 익숙하지 않은 것으로 추측되어 큰 차이가 없는 것으로 판단된다.

질문2 (제시된 설명인터페이스가 추천상품 구매여

부에 어느 정도 도움이 되었습니까?)에 대한 평가 점수의 평균과 표준편차는 다음의 〈표 10〉과 같다. 결과를 보면 〈표 9〉의 설명 인터페이스의 이해 정도와 〈표 10〉의 도움이 된 정도의 순위는 상당히 유사하다는 것을 알 수 있다. 이것은 설명인터페이스가 쉽고, 이해도가 높을수록 구매의사 결정에 기여도가 높아진다는 것을 뜻한다. 특히 1~4위 정보를 보면 유사한 선호도를 가진 고객의 선호도 정보, 구매 정보는 고객이 의사결정을 하는데 상당히 긍정적인 영향을 미친다는 것을 알 수 있다. 또한 실질적인 구매와 관련된 할인율과 같은 마케팅 정보 또한 구매 의사결정에 중요한 요소로 작용한다고 볼 수 있다. 20개의 설명인터페이스 중 19개가 평균점수 3점 이상이라는 것은 설명기능이 추가됨으로써 구매의사를 결정하는데 도움이 되었다는 것을 알 수 있다. 하지만 실험환경을 고려하면 상위 3개의 설명인터페이스가 구매의사결정에 좀더 높은 효과를 줄 수 있으며, 상위 9위까지의 설명인터페이스가 향후 연구에 따라 추천에 관한 의사결정 효과를 향상시키는데 도움을 줄 수 있을 것으로 판단할 수 있다. 그리고, 20번과 같은 모호한 표현으로 명확한 설명을 하지 못하는 설명기능은 오히려 구매 의사결정에 부정적인 영향을 준다는 것을 알 수 있다.

또한 어떤 유형의 설명인터페이스가 사용자 구매 의도에 도움을 줄 것인가를 통계적으로 검증하였다. 〈표 11〉은 분산분석을 실시한 결과를 보여준다. F

〈표 10〉 설명인터페이스가 구매의도에 도움을 준 정도

	설명인터페이스	평균	표준편차
1	이웃의 선호도(5개의 별)	3.710	1.038
2	이 상품은 특가기획상품으로 행사기간 중에 정가의 40%가 할인 됩니다(기술)	3.609	1.060
3	구매 비율(그래프)	3.536	1.069
4	이 상품을 추천 받은 고객 중 64%가 구매 하였습니다.	3.371	1.014
5	이 상품은 지난 주에 가장 많이 구매한 상품으로 이주의 상품으로 선정 되었습니다(기술)	3.277	1.031
6	상품조회 비율(그래프)	3.261	0.993
7	이 상품은 과거 고객님의 구매한 상품들과 유사합니다(기술)	3.253	1.229
8	장바구니 담은 비율(그래프)	3.248	1.036
9	장바구니 담은 비율(기술)	3.235	0.947
10	이 상품은 고객님의 유사한 선호도를 가진 30명의 구매 정보를 기반으로 추천 되었습니다(기술)	3.185	1.108
11	고객님의 상품 분류별 선호도(그래프)	3.181	1.251
12	장바구니 담은 비율(테이블)	3.180	0.989
13	상품조회 비율(기술)	3.162	0.972
14	상품조회 비율(테이블)	3.161	0.926
15	고객님의 상품 분류별 선호도(기술)	3.149	1.218
16	고객님의 상품 분류별 선호도(테이블)	3.138	1.058
17	구매 비율(기술)	3.090	1.111
18	높은 정확성이 검증된 추천 시스템은 이 상품이 고객님의 적합할 것이라고 예측했습니다(기술)	3.058	1.120
19	구매 비율(테이블)	3.049	1.206
20	이 상품은 주요 일간지와 유명잡지에 여러번 소개 되었습니다(기술)	2.825	1.250

값은 3.160이며, P값은 0.000으로 유의수준 5% 내에서 매우 유의하다고 할 수 있다. 그러므로 설명 인터페이스는 유형에 따라 사용자 구매의도에 영향

을 준다고 할 수 있다.

추가로 사후분석을 한 결과, 설명유형 18번과 설명유형 2, 3, 6, 8, 9, 10번과 통계적으로 유의한

〈표 11〉 설명인터페이스 유형에 따른 구매의도 분석결과

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	74.606	19	3.927	3.160	.000
Within Groups	2295.169	1847	1.243		
Total	2369.775	1866			

차이가 있고, 설명유형 19번과 설명유형 2, 3, 8, 9, 10번도 통계적으로 유의한 차이가 있는 것으로 나타났지만, 나머지 설명유형들 간의 차이는 통계적으로 없는 것으로 나타났다. 이러한 결과로 볼 때, 피실험자들에게 구매에 도움을 주는 정보는 이해정도와 유사하게 이웃의 선호도 (별표시) 정보와 특정상품에 대한 행사정보와 같은 마케팅 정보가 사용자 구매의도에 영향을 미치는 것으로 나타났다. 하지만 다른 형태의 설명유형은 피실험자들에게 익숙하지 않은 것으로 나타났으며, 설명유형을 구현하는데 있어서 좀 더 세심한 실험이 요구된다고 사료된다.

VI. 결 론

본 연구는 기존 협업필터링 기반 상품추천시스템의 문제점을 해결하기 위해 상품을 추천 받은 고객이 구매의사를 결정 할 때 도움을 주는 설명기능에 관한 연구이다. 본 연구의 의의를 정리하면 다음과 같다. 첫째, 기존에 제안한 WebCF-PT추천방법론에 설명기능을 추가한 WebCF-Exp추천시스템의 프로토타입을 구현하였다. 둘째, 추천시스템의 성능평가를 위해 가상의 EBIB Research 쇼핑몰을 ASP, MSSQL, VB6.0을 이용하여 온라인 상으로 구현하여 실험에 필요한 실제 데이터를 수집하였다. 셋째, 피실험자들에게 가상쇼핑몰을 이용하게 하고, 상품을 추천한 다음 피실험자들이 구매한 데이터를 분석한 결과, WebCF-Exp추천시스템의 성능이 우수하다는 것을 입증하였다. 넷째, 설문조사를 통하여 WebCF-Exp추천시스템에 효과가 높은 설명인터페이스 유형을 알아보았다. 설명인터페이스 중 이

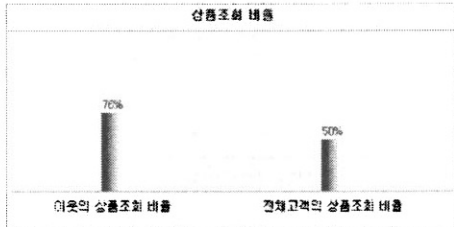
웃들의 선호도 정보와 평가정보, 마케팅 정보가 높은 점수를 얻었고, 이런 정보들을 그래프나 별표시 등의 시각적인 유형으로 설명인터페이스를 구현하는 것이 고객의 이해와 구매를 돕는데 더 유용한 것으로 나타났다. 다섯째, 설명 인터페이스의 이해도가 높을수록 고객의 구매의사결정에 도움이 된다는 것을 알 수 있었다. 결론적으로, 고객이 이해하기 쉬운 설명기능을 협업필터링기반 추천시스템에 추가하면 고객들의 구매의사 결정에 도움을 준다는 것이 입증되었다.

본 연구의 한계점은 다음과 같으며 추후연구를 통해 개선되어야 할 점들이다. 첫째, 2달이라는 짧은 기간에 이루어진 실험으로 인하여 충분한 정보를 획득하지 못한 문제점이 있다. 둘째, 고객의 실제 구매비용이 지출된 것이 아니라 무료의 사이버 머니를 사용함으로써 상품 선택에 따른 신중성과 정확성에 대한 문제점이 지적될 수 있다. 셋째, EBIB Research 쇼핑몰의 상품도메인이 화장품 하나의 상품군으로만 구성되어 있으므로, 다양한 상품도메인에 관한 정보를 반영하지 못한다는 문제점이 있을 것으로 생각된다. 넷째, 실제 상품구매에 영향을 미치는 요인으로는 상품조회비용, 장바구니 담은 비율, 구매비용, 유명매체의 추천 등 추천 컨텐츠와 관련된 내용이 영향을 미칠 것으로 판단되어 이에 대한 분석이 더 필요할 것으로 판단된다. 마지막으로 설명기능을 제공 받은 후 고객의 피드백에 관한 것이 고려되지 않았으므로 추후 이를 고려한 연구가 필요할 것으로 생각된다.

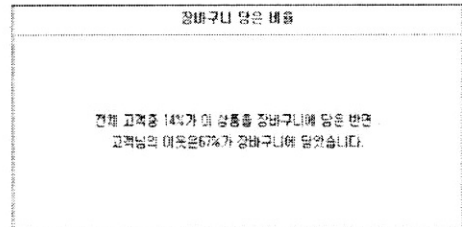
참고문헌

- 김재경, 안도현, 조윤희 (2003), "Development of a Personalized Recommendation Procedure Based on Data Mining Techniques for Internet Shopping Malls," *한국지능정보시스템학회*, 9, 177-191.
- 김재경, 안도현, 이희애 (2004), "웹마이닝과 상품계층도를 이용한 협업필터링 기반 개인별 상품추천시스템: WebCF-PT," *한국지능정보시스템학회 2004 춘계학술대회*.
- Billsus, D., and M. J. Pazzani (1998), "Learning collaborative information filters," *In Proc. 15th International Conference on Machine Learning*, 46-45.
- Cho, Y. H., J. K. Kim, and S. H. Kim (2002), "A personalized recommender system based on Web usage mining and decision tree induction," *Expert Systems with Applications*, 23, 329-342.
- Cho, Y. H. and J. K. Kim (2004), "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, 26, 233-24.
- Dahlen, B. J., J. A. Konstan, J. L. Herlocker, N. Good, A. Borchers, and J. Riedl (1998), "Jump-starting movielens: User benefits of starting a collaborative filtering system with dead data. University of Minnesota TR 98-017.
- Han, J. and Y. Fu (1999), "Mining Multiple-Level Association Rules in Large Databases," *IEEE Transaction on Knowledge and Data Engineering* 11, 798-804.
- Herlocker, J. L., J. A. Konstan, A. Borchers, and J. Riedl (1999), "An Algorithmic Framework for Performing Collaborative Filtering," *In Proc. Conference on Research and Development in Information Retrieval*, 230-237.
- Herlocker, J. L., J. A. Konstan, and J. Riedl (2000), "Explaining Collaborative Filtering Recommendations," *In Proc. 2000 ACM Conference on Computer Supported Cooperative Work*, 241-250.
- Hill, W., L. Stead, M. Rosenstein, and G. W. Furnas (1995), "Recommending and Evaluating Choices in a Virtual Community of Use," *In Proc. Human Factors in Computing Systems*, 194-201.
- Kim, S. H. and J. K. Kim (1991), "Explanation in a Decision-Theoretic Consulting System: An Axiomatic Approach," *International Journal of Applied Artificial Intelligence*, 5, 393-409.
- Lawrence, R. D., G. S. Almasi, V. Kotlyar, M. S. Viveros, and S. S. Duri (2001), "Personalization of supermarket product recommendations," *Data Mining and Knowledge Discovery*, 5, 11-32.
- Mobasher, B., R. Cooley, and J. Srivastava (2000), "Automatic Personalization based on Web usage mining," *Communications of the ACM*, 43, 142-151.
- Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl (1994), "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *In Proc. Computer Supported Cooperative Work, Chapel Hill, NC*, 175-186.
- Sarwar, B., G. Karypis, J. A. Konstan, and J. Riedl (2000), "Analysis of recommendation algorithms for e-commerce," *In Proc. ACM E-Commerce*, 158-167.

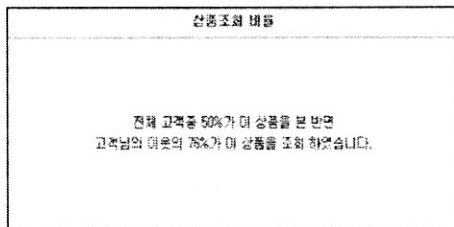
〈부록: 설명인터페이스〉



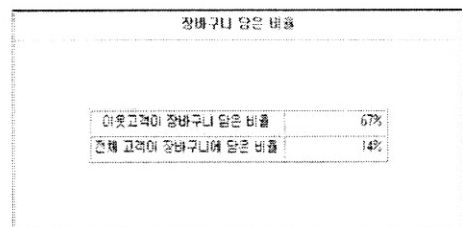
〈설명 1〉 상품조회 비율



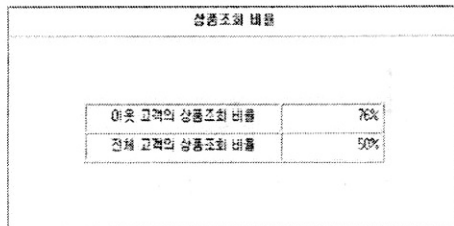
〈설명 5〉 장바구니 담은 비율



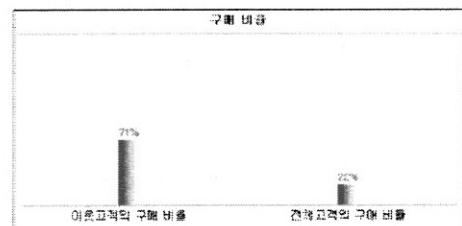
〈설명 2〉 상품조회 비율



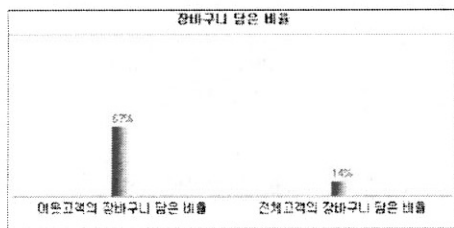
〈설명 6〉 장바구니 담은 비율



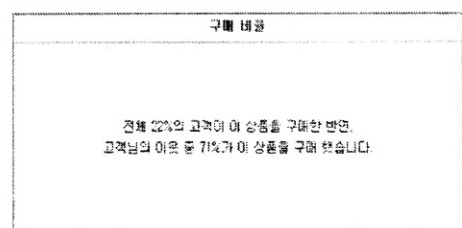
〈설명 3〉 상품조회 비율



〈설명 7〉 구매 비율



〈설명 4〉 장바구니 담은 비율



〈설명 8〉 구매 비율

구매 비율	
이웃고객의 구매 비율	71%
전체고객의 구매 비율	22%

〈설명 9〉 구매 비율

이 상품은 고객님과 유사한 선호도를 가진
20명의 구매정보를 기반으로 추천 되었습니다.

〈설명 13〉 이웃의 수

이 상품은 주요 일간지와 유명 잡지에 의해 여러번 소개 되었습니다.

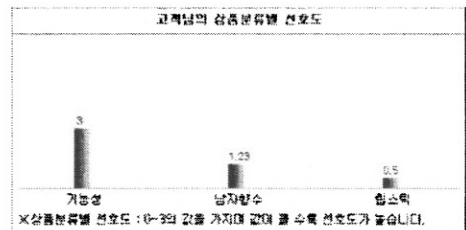
〈설명 10〉 유명 매체의 추천

이 상품을 추천한 고객중 5%가 구매 했습니다.

〈설명 14〉 이웃의 구매 비율

높은 정확성에 검증된 추천 시스템은
이 상품이 고객님께 적합할 것이라고 예측 했습니다.

〈설명 11〉 추천시스템의 예측



〈설명 15〉 목표고객의 상품군별 선호도

이 상품은 과거 고객님이 구매한 상품들과 유사합니다.

〈설명 12〉 구매한 상품과의 유사 정도

고객님의 상품군별 선호도는
기능성 1.23 남자합수 1.23 합스택 0.5입니다.
※상품군별 선호도 : 0-3의 값을 가지며 값이 클 수록 선호도가 높습니다.

〈설명 16〉 목표고객의 상품군별 선호도

설명기능을 추가한 협업필터링 기반 개인별 상품추천시스템: WebCF-Exp

고객의 상품분류별 선호도	
가방형	3%
남자형	1.23%
립스틱	0.5%

*상품분류별 선호도와 3의 값을 가지면 같이 볼 수록 선호도가 높습니다.

〈설명 17〉 목표고객의 상품군별 선호도

이 상품은 특가기획상품으로
행사기간 중에 정가의 40%가 할인 됩니다.

〈설명 19〉 마케팅 정보1.

이웃의 선호도: ★★★★★

〈설명 18〉 이웃의 선호도(5개의 별)

이 상품은 지난 주에 가장 많이 구매한 상품으로
이주의 상품으로 선정 되었습니다.

〈설명 20〉 마케팅 정보2.

An Explanation based Recommender System using Collaborative Filtering : WebCF-Exp

Jae Kyeong Kim* · Hee Ae Lee** · Do Hyun Ahn*** · Yoon Ho Cho****

Abstract

The continuous growth of the Internet and e-commerce has allowed companies to provide customers with more choices on products. Increasing choice has also caused product overload where the customer is no longer able to effectively choose the products he/she is exposed to. A promising technology to overcome the product overload problem is recommender systems that help customers find the products they would like to purchase. To date, a variety of recommendation techniques have been developed. Collaborative Filtering (CF) is the most successful recommendation technique, which has been used in a number of different applications such as recommending movies, articles, books, Web pages, etc. However, its widespread use has exposed some limitations, such as sparsity, scalability, and black box.

Many researchers have focused on sparsity and scalability problems but a little has tried to solve the black box problem. Most CF recommender systems are black boxes, providing no transparency into the working of the recommendation. To overcome the black box problem, it is developed a recommender system named WebCF-Exp (Web usage mining driven Collaborative Filtering with Explanation facilities).

Explanation facilities make it possible to expose the reasoning and data behind a recommendation. Therefore, they provide us with a mechanism for handling errors that come with a recommendation. Furthermore, it is proposed to use web usage mining and product taxonomy to enhance the recommendation quality for e-commerce environment. Web usage mining populates the rating database by tracking customers' shopping behaviors in the Web.

* Professor, School of Business Administration, Kyung Hee University.

** Researcher, eBusiness Information Center, Kyung Hee University.

*** Doctoral Candidate, Graduate School of Business Administration, Kyung Hee University.

**** Associate Professor, School of e-Business, Kookmin University.

thereby leading to better recommendations. The product taxonomy is used to improve the performance through dimensionality reduction of the rating database.

The overall procedure of WebCF-Exp consists of two phases: recommendation phase and explanation phase. The recommendation phase is divided into four sub-phases: grain specification, customer profile creation, neighborhood formation, and recommendation generation. The explanation phase consists of white box model and black box model. A white box model is one way to build explanation interfaces using detailed process or data such as neighbor ratings, the ratio of click, and the ratio of purchase. Black box model is the other way of which there is no detailed process or data available. In black box model, we focus on ways to justify recommendation that are independent of the process such as marketing or event information.

WebCF-Exp recommender system is operated by four agents: Web log analysis agent, Data transformation agent, Recommendation agent, and Explanation agent. Web log analysis agent manages Web log database through periodic collecting, parsing and analyzing Web server log files such as access logs, referrer logs, agent logs and cookie files. Thus, the users can easily access and analyze them like other operation databases. Data transformation agent creates and manages the data mart that provides data indispensable to accomplish recommendation tasks. Recommendation agent makes a personalized recommendation list for each target customer. Explanation agent provides interfaces which expose the reasoning and data behind a recommendation. Twenty different explanation interfaces are developed as white box model and black box model.

To test the performance of WebCF-Exp, it is developed a prototype internet shopping mall named EBIB (e-Business & Intelligent Business) and interactive interfaces. Experiments are conducted with the data provided by EBIB Research Internet shopping mall. Our experimental result shows that WebCF-Exp recommendation system shows better performance than the CF recommendation system without explanation facilities. And explanation types of five stars, simple graph, and showing the evaluation results of similar customers, show better performance than other types. Furthermore, as customers understand explanation interfaces better, it results that customers purchase more products.

Key words: Collaborative Filtering, Recommendation system, Explanation, Internet shopping mall.