

# 개인별 상품추천시스템, WebCF-PT: 웹마이닝과 상품계층도를 이용한 협업필터링

김재경\*, 안도현\*, 조윤호\*\*

## A Personalized Recommender System, WebCF-PT: A Collaborative Filtering using Web Mining and Product Taxonomy

Jae Kyeong Kim, Do Hyun Ahn, Yoon Ho Cho

Recommender systems are a personalized information filtering technology to help customers find the products they would like to purchase. Collaborative filtering is known to be the most successful recommendation technology, but its widespread use has exposed some problems such as sparsity and scalability in the e-business environment. In this paper, we propose a recommendation system, WebCF-PT based on Web usage mining and product taxonomy to enhance the recommendation quality and the system performance of traditional CF-based recommender systems. Web usage mining populates the rating database by tracking customers' shopping behaviors on the Web, so leading to better quality recommendations. The product taxonomy is used to improve the performance of searching for nearest neighbors through dimensionality reduction of the rating database. A prototype recommendation system, WebCF-PT is developed and internet shopping mall, EBIB(e-Business & Intelligence Business) is constructed to test the WebCF-PT system.

**Keywords :** Collaborative Filtering, Recommendation System, Web Usage Mining, Product Taxonomy, Internet Shopping Mall

---

\* 경희대학교 경영대학 e비즈니스 전공

\*\* 국민대학교 e비즈니스 학부(교신저자)

## I. 서론

인터넷의 급속한 성장과 발달로 이를 기반으로 한 전자상거래가 빠르게 증가하고 있다. 인터넷 쇼핑몰이나, 콘텐츠 제공업체(Content Provider)와 같은 기업들은 전자상거래의 급성장으로 인하여 생존우위와 정보과부하 현상을 해결하기 위한 새로운 마케팅 전략이 필요하게 되었다. 또한 고객들은 온라인기업이 증가함에 따라 다양하고, 많은 상품을 선택할 수 있는 기회가 주어지지만 본인의 요구에 가장 적합한 상품을 찾기 위해서 더 많은 정보를 처리해야 하는 부담이 생기게 되었다. 이와 같은 이유로 기업은 고객별로 차별화된 일대일 마케팅(One-To-One Marketing)과 고객의 입장에서 고객을 이해하고 고객과의 관계를 강화시켜 나가는 CRM(Customer Relationship Management, 고객관계관리) 전략 등을 사용함으로써 기업의 경쟁력 강화에 주력하고 있다. 특히 CRM의 여러 분야 중에서 구매촉진(Campaign Management)을 위해 추천시스템이 많이 사용되고 있다. 이러한 추천시스템(Recommendation System)은 통계적 기법과 지식 탐사기술(Knowledge Discovery Technology)을 이용하여 고객의 요구에 가장 부합되는 상품을 추천해주는 시스템으로서, 고객들의 편의를 도모하고 교차판매(cross selling) 및 매출 증대에 초점을 맞춘 시스템이다[김재경 외, 2003; Sarwar et al., 2000a].

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. 이러한 추천시스템의 핵심은 추천알고리즘에 있으며, 현재까지 가장 선호되고 있는 추천알고리즘은 협업필터링(Collaborative Filtering: CF)으로 영화, 웹사이트, 상품, 뉴스 등 여러 분야에서 사용되고 있다[Basu et al., 1998; Billsus and Pazzani, 1998;

Cho and Kim, 2004; Hill et al., 1995; Kim et al., 2002; Lawrence et al., 2001; Resnick et al., 1994]. 협업필터링 기법은 목표 고객과 유사한 선호도를 보이는 이웃고객들이 구매한 상품들 중 구매할 가능성이 가장 높은 상품을 추천하는 기법이다. 그러나 이러한 협업필터링 기법은 다른 알고리즘보다 우수함에도 불구하고 다음과 같은 문제점들이 존재한다[Cho and Kim, 2004; Kim et al., 2002; Melville et al., 2001; Sarwar B., 2001]. 첫째, 입력 데이터의 희박성(sparsity) 문제이다. 웹사이트에서 판매되는 상품의 수가 기하급수적으로 증가함에 따라 고객의 선호도가 입력되지 않은 상품의 개수가 상대적으로 많아짐으로 인해 이웃 고객군을 형성하는 과정에서 매우 적은 수의 평가 데이터만을 사용함으로써 유사도 측정에 신뢰성이 떨어지고, 이는 결국 상품 추천의 질을 떨어뜨리는 요인으로 작용한다. 둘째, 시스템의 확장성(scalability)문제이다. 고객과 상품의 수가 증가함에 따라 이웃 고객군을 찾기 위한 연산량은 기하급수적으로 늘어날 수밖에 없기 때문에 실시간으로 추천을 목적으로 하는 상품추천시스템에서는 심각한 시스템 확장성문제에 직면하게 된다.

따라서 본 연구에서는 이와 같은 협업필터링의 문제점들을 해결하기 위해 웹마이닝(Web mining)과 상품계층도(Product Taxonomy)를 이용한 협업필터링 기반 상품추천시스템, WebCF-PT(Web-based Collaborative Filtering using Product Taxonomy)를 개발하고자 한다. 웹마이닝은 입력 데이터의 희박성문제를 해결하기 위해 고객이 인터넷 쇼핑몰에서 상품을 처음 인식하는 행동에서부터 상품을 클릭하고 장바구니에 담아 상품을 구매하기까지의 과정에 대한 웹기반 정보를 이용하여 보다 많은 고객 선호도 정보를 확보하기 위해 적용한 것이며, 상품계층도는 관련 있는 상품들을 특정 상품군으로 군집화 하여 선호도 입력 데이터의 차원을 축소함으로써 입력 데이터의 희박성과 시스템 확장

성문제를 동시에 해결하기 위해 적용한다. 또한 WebCF-PT시스템을 적용하기 위하여 실험목적의 인터넷 쇼핑몰을 개발하여, 실제로 적용되는 사례를 제시하였다.

본 연구의 기여도를 살펴보면, 첫째, 상품계층도를 입력 데이터의 희박성을 줄이고, 시스템 확장성을 향상시키기 위해 적용하였다. 둘째, 고객의 선호도를 파악하기 위해 기존의 일반적인 방법인 고객들에게 직접 설문을 통하여 파악하지 않고 암묵적으로 고객의 온라인 쇼핑 행태를 분석하는 웹마이닝 기법을 개발하였다. 셋째, WebCF-PT 프로토타입 시스템을 개발하고 이를 적용하기 위하여 실험 목적의 EBIB Research 쇼핑몰을 구축하였다.

본 연구의 구성을 살펴보면 다음과 같다. II장에서는 기존 연구에 대해 살펴보고, III장에서는 본 연구에서 제안하는 웹마이닝과 상품계층도를 활용한 협업필터링 기반 추천방법론에 대해 설명한다. IV장에서는 제안한 방법론으로 구현된 WebCF-PT 추천시스템을 설명하고, V장에서는 WebCF-PT시스템을 적용하기 위하여 개발한 EBIB Research 쇼핑몰을 소개하며, VI장에서는 연구의 결론을 제시한다.

## II. 관련 연구

### 2.1 상품추천시스템

협업필터링(Collaborative Filtering: CF)은 웹을 기반으로 하는 전자쇼핑몰에서 이용되고 있는 성공적인 상품추천기법중의 하나로써, 목표고객과 유사한 구매이력을 보이는 이웃 고객들의 상품에 대한 선호를 바탕으로 목표고객에게 유용한 상품을 추천하는 방법이다[김재경 외, 2002; Resnick et al., 1994; Sarwar et al., 2000]. 일반적으로 이러한 협업필터링 기반 상품 추천 과정은 크게 입력 데이터 구성, 이웃 집단 탐색, 추천 상품 결정 단계로 나뉘볼 수 있으며, 이러

한 과정을 자세히 살펴보면 다음과 같다.

- (1) 입력 데이터 구성(Data Representation): 협업필터링 기반 상품추천시스템에서의 입력 데이터는 보통  $n$ 개의 상품에 대하여  $m$ 명 고객의 구매 트랜잭션의 집합으로 구성되며, 보통  $n \times m$ 의 고객-상품 행렬  $R$ 로 표현될 수 있다.
- (2) 이웃 집단 탐색(Neighborhood Formation): 고객간의 유사도를 계산하여 이웃 집단을 탐색하는 과정이다. 두 고객의 유사도를 측정하는 방법으로써 피어슨 상관계수(Pearson Correlation), 코사인(Cosine) 등을 사용한다.
- (3) 추천 상품 결정(Generation of Recommendation): 상품 추천을 위한 마지막 단계로서 설정된 이웃 집단으로부터 상위  $N$ 개의 추천 상품 목록을 이끌어 내는 단계이다.

그러나 이러한 협업필터링 기반 추천시스템은 인터넷 쇼핑몰의 상품과 고객수의 급속한 증가로 인해 입력 데이터의 희박성과 시스템 확장성문제를 노출시키고 있다[Billsus and Pazzani, 1998; Sarwar et al., 2000a].

### 2.2 웹마이닝

인터넷 쇼핑몰에 있어서 획득 가능한 고객정보에는 오프라인 기업에서와 같은 구매이력 외에도 고객들이 쇼핑몰 사이트를 방문할 때 발생하는 웹로그 정보가 있다. 이 정보는 인터넷 기업에 있어서 고객분석을 할 때 간과해서는 안될 매우 중요한 정보이며, 특히 상품추천의 경우, 고객의 성향이나 상품선호도 분석에 있어 핵심적인 기초정보로 활용할 수 있다. 따라서 본 연구에서는 이러한 웹로그 정보를 적극적으로 상품추천에 활용하기 위해 웹마이닝 기법을 이용한다. 웹마이닝(Web Mining)은 웹사이트에서 고객이동경로 즉 클릭스트림 정보를 가지고 있

는 웹로그로부터 고객들의 접속관계·패턴·규칙 등을 찾아내고, 모형화해 유용한 마케팅 정보로 변환시키는 일련의 과정을 말한다[Mobasher et al., 2000]. 웹마이닝의 전체 프로세스는 크게 데이터 전처리와 패턴분석과정으로 분류할 수 있다. 데이터 전처리는 획득한 웹로그 데이터를 분석에 적합한 형태로 정제하여 불필요한 데이터를 제거함으로써 원천데이터의 용량을 감소시킬 수 있다[Cooley et al., 1999]. 두 번째, 패턴분석과정은 연관성규칙, 연속적인 패턴발견 등을 이용하여 고객의 쇼핑행위에서 구매패턴을 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 희박성문제를 해결할 수 있을 것으로 기대된다.

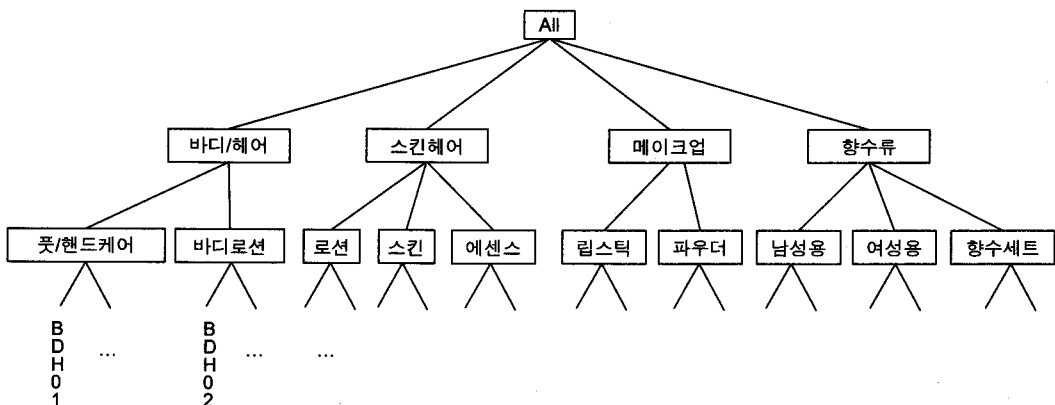
### 2.3 상품계층도

상품계층도는 개별상품을 추상 개념이 낮은 상품 클래스로 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 보여주고 있다. 최근 많은 데이터 마이닝 관련 연구

들이 데이터 분석에서의 상품 계층도의 필요성 및 중요성을 지적하고 있다[Han and Fu, 1999; Lawrence, et al., 2001]. 상품계층도상에서 유사한 선호도 패턴을 갖는 개별상품들을 특정 상품군으로 군집화하여 입력 데이터의 차원을 축소하면 이웃 집단 탐사 과정에서 계산속도의 향상을 도모하여 시스템 확장성문제를 해결할 수 있을 것이다.

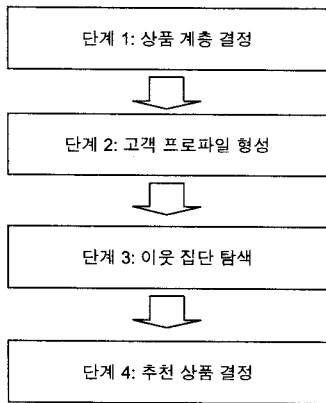
### Ⅲ. WebCF-PT를 이용한 개인별 상품추천 방법

본 연구에서는 기존 협업필터링의 문제점들을 해결하기 위해 웹마이닝과 상품계층도를 이용한 협업필터링 기반 상품추천방법(Web-based Collaborative Filtering using Product Taxonomy: WebCF-PT)을 소개한다. <그림 2>는 WebCF-PT추천방법의 개략적 절차를 나타내며, 첫 단계는 '상품 계층 결정' 단계로 데이터베이스 내의 모든 상품을 마케팅 전문가가 특정그룹으로 분류·재구성하여 입력 데이터의 차원을 감소시키는 단계이다. 두 번째 단계는 '고객 프로필 형성' 단계로 인터넷 쇼핑몰에서 개별고객의 쇼핑행위를 추적하여 얻은 데이터를 통해 목표고객의 상품 선호도 정보를 발견하고, 분석한 정보를 이용하여 고객 프로필을 형성하는



<그림 1> 상품계층도

단계이다. 세 번째 단계는 ‘이웃 집단 탐색’ 단계로 형성된 고객 프로파일을 이용하여 고객간의 유사도를 계산하고 이를 기반으로 목표고객과 유사한 성향을 가진 고객들을 선택하여 이웃을 형성하는 단계이다. 마지막 단계는 ‘추천 상품 결정’ 단계로 형성된 이웃들의 쇼핑행위를 기반으로 선호도가 높은 상위 N개의 상품을 선택하여 추천 상품의 목록을 결정하는 단계이다.



<그림 2> WebCF-PT추천방법의 개략적 절차

각각의 단계에서 수행되는 기능들을 단계별로 자세히 설명하면 다음과 같다.

### 3.1 단계 1: 상품 계층 결정

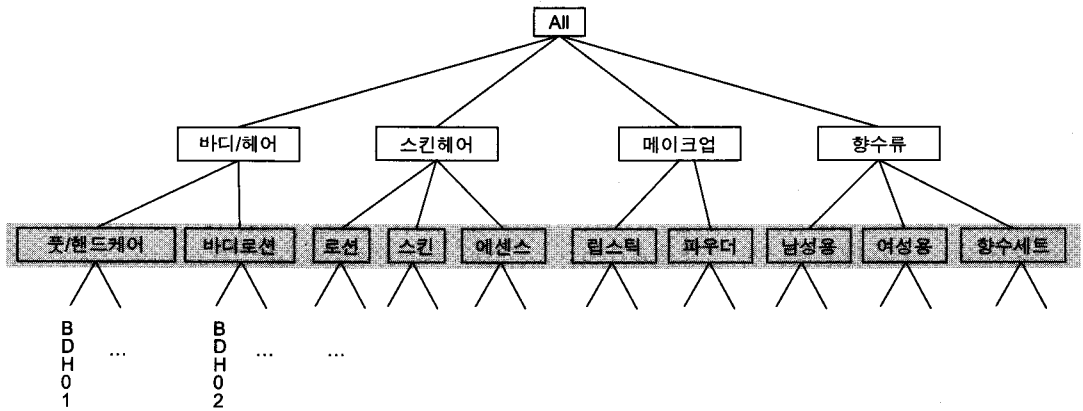
상품 계층 결정 단계에서는 상품계층도를 이용하여 유사한 상품들을 정의하고 그룹화함으로써 입력 데이터의 차원을 감소시키는 단계이다. 이러한 상품계층도는 상품 구매 빈도를 기반으로 마케팅 전문가에 의해 유사한 상품들을 정의하고 그룹화함으로써 상품 계층을 결정하게 된다. 상품 계층을 결정하는데 좀더 유연한 방법을 채택하기 위해 Cho and Kim[2004]이 제안한 그레인(Grain) 개념을 도입한다. 그레인은 상품계층도에서 개별상품의 개수가 너무 많을 때 개별 상품들을 클래스로 묶을 때 사용하기 위한 방법으로 Adomavicius and Tuzhilin[2001]에 의해

소개된 “컷(cut)”의 기본원리를 이용한 것이다. 그레인  $G$ 는 상품계층도에서 루트노드(root node)를 제외한 모든 노드의 부분집합으로 개별상품에서부터 루트노드까지의 경로를 말한다. 그러므로 모든 개별 리프노드(leaf node)는 그것에 대응하는 그레인노드를 가지고 있으며, 이를 “그레인 상품 클래스(grain product class)”라고 부른다. 즉 이러한 그레인 상품 클래스가 상품 계층이 된다. 그레인  $G$ 가 주어졌을 때, 상품  $x$ 에 대응하는 그레인 상품 클래스  $class_G(x)$ 는 다음 식 (1)과 같이 정의한다. 리프노드는 개별상품 노드이며, 그레인 노드는 상품 구매 빈도를 기반으로 마케팅 전문가에 의해 유사한 상품들을 정의하고 그룹화 하여 결정된다. 예를 들면, <그림 3>(c)혼합 수준 그레인 예를 가지고 설명하면, 마케팅 전문가에 의해 구성된 상품계층도에서 상품 계층 결정, 즉 그레인을 결정하기 위해 상품 구매 빈도를 파악한 결과 립스틱과 파워더의 빈도가 현저히 낮기 때문에 한 수준 높은 메이크업을 분석대상으로 한다. 즉, 상품의 판매 현황 등을 고려하여 메이크업과 향수류는 에센스, 스킨 등의 개별상품과 같은 분석대상이 되는 것이다. 물론 여기서 보여주는 상품계층도는 간단한 예를 보여주고 있지만 실제적으로 상품구매 빈도를 기반으로 상품계층, 즉 그레인 노드가 결정된다. 자세한 내용은 Cho and Kim [2004]을 참조하기 바란다.

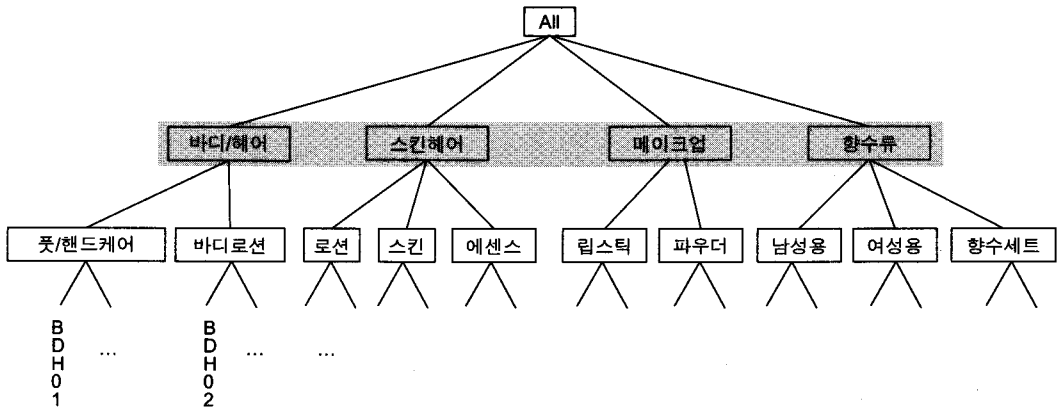
$$class_G(x) = \begin{cases} x, & \text{if } x \in G, \\ class_G(\text{parent}(x)), & \text{otherwise} \end{cases} \quad (1)$$

<그림 3>은 여러가지 서로 다른 그레인의 예를 보여주고 있다. 각 상품계층도의 그림자 부분이 그레인이다.

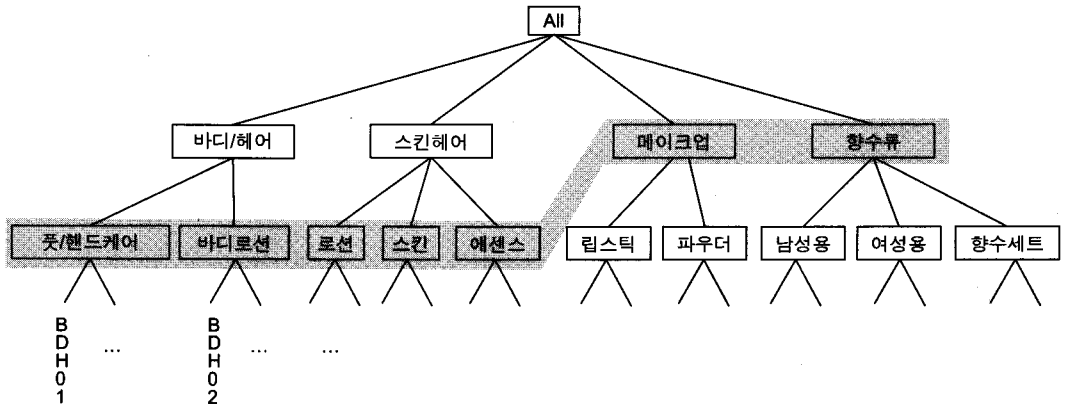
다음의 상품계층도를 이용하여 본 연구에서 제안한 그레인 상품 클래스를 정의해 보면 <그림 3>(a)는 가장 낮은 수준의 그레인  $G = \{\text{풋}/$



(a) 낮은 수준의 그레인(lower level grain)



(b) 높은 수준의 그레인(higher level grain)



(c) 혼합 수준 그레인(level-crossing grain)

<그림 3> 상품계층도에서 그레인의 예

핸드케어, 바디로션, 로션, 스킨, 에센스, 립스틱, 파우더, 남성용, 여성용, 향수세트), <그림 3>(b)는 한 수준 높은 그레인  $G = \{\text{바디/헤어, 스킨케어, 메이크업, 향수류}\}$ 로 분류 할 수 있다. 이와 같이 판매 빈도가 비슷하다면 그레인은 <그림 3>(a)와 <그림 3>(b)와 같이 동일 수준에 있는 노드들의 집합일 수도 있다. 또한 판매 빈도와 마케팅 전문가의 지식에 따라 유연하게 상품 클래스가 달라질 수도 있다. 만약 풋/핸드 케어 상품 클래스에 속해 있는 BDH01 등의 개별상품들이 판매 빈도가 현저히 낮을 경우, 상품계층도상에서 한 수준 더 높은 수준을 선택할 수도 있다. <그림 3>(a)에서  $class_G(BDH01) = \text{'풋/핸드 케어'}$ 와  $class_G(BDH02) = \text{'바디로션'}$ 이지만, <그림 3>(b)에서  $class_G(BDH01) = \text{'바디/헤어'}$ 이고 또한  $class_G(BDH02) = \text{'바디/헤어'}$ 이다. 그러므로 상품 'BDH01'과 'BDH02'는 가장 낮은 수준에서는 개별 상품이지만 그레인의 수준에 따라 같은 상품군이 될 수도 있다. 그레인은 <그림 3>(c)에서처럼 다른 레벨의 상품군과 혼합하여 형성할 수도 있는데, 그것은 추천하는 상품의 상대적인 중요도나, 상품판매량 등에 따라 결정할 수 있다. <그림 3>의 예는 입력 데이터의 차원 감소 관점에서 개별상품들의 판매 빈도가 낮아서 한 수준 위의 계층을 선택하고 있다. 예를 들면, 그레인을 형성할 때, 자주 팔리는 상품은 상품계층도의 낮은 수준에 머물러 있게 하고, 구매 빈도가 매우 드문 상품의 경우에는 그러한 상품들을 클래스로 취급하여 상품계층도의 높은 수준에 있게 한다. 혼합수준 그레인은 개별 상품의 판매량에 따라 다양한 상품계층도의 수준으로 그레인을 형성함에 따라 결과적으로 그레인의 개별 단위의 판매가 비슷하게 발생되도록 한다. 상위 레벨의 그레인은 개별상품보다 좀더 거시적인 상품클래스를 거래 데이터로 다루게 되며, 개별상품 판매가 적을 경우에 적절하다. 또한 이렇게 상품계층도상에서 상품들의 판매량에 따라 특정 상품계층으

로 근집화하여 입력 데이터의 차원을 축소하면 이웃 집단 탐색 단계에서 계산 속도의 향상을 도모할 수 있다. 따라서 적절한 그레인 기반의 상품추천은 협업필터링의 문제점인 희박성과 확장성문제를 해결 할 수 있게 된다. 하지만 그레인 기반의 상품 추천 시 바디/헤어를 사용한다는 점에서 두 사람이 유사하다고 본다 해도 실제로 한 명은 '풋/핸드 케어'만을 다른 한 명은 '바디로션'만을 쓴다면 추천의 정확성에 문제가 발생할 수도 있다는 문제가 제기될 수 있다. 하지만 상품계층도를 구성할 때 기본 전제가 상품의 유사성을 고려하여 마케팅 전문가에 의해 구성되며, 만약 실제 사용자들이 개별 상품에 대한 빈도가 많을 경우 그 수준에서 계층 즉 그레인이 결정되며, 상품 판매 빈도가 낮을 경우에만 유연하게 상위 수준으로 그레인이 결정되기 때문에 시스템 확장성 및 희박성 문제를 어느정도 해결할 수 있다고 본다. 또한 다음 단계에서 웹마이닝을 통하여 희박성 문제를 보완하기 때문에 추천의 정확성에도 도움이 된다고 생각된다.

### 3.2 단계 2: 고객 프로파일 형성

고객 프로파일 형성은 상품에 대한 고객의 선호도를 나타낸 것으로 협업필터링 기반의 추천 방법에 있어서 추천의 정확도와 관련해서 매우 중요한 단계이다. 기존 협업필터링 기반 추천방법에서는 고객의 구매 정보만을 가지고 고객 프로파일을 형성함으로써 인해 입력 데이터 희박성 문제를 발생시켰다. 따라서 이 연구에서는 기존 협업필터링 기반 상품추천방법과 다르게 2.2절에서 설명한 것과 같이 웹마이닝을 이용하여 클릭스트림 데이터로부터 고객의 상세한 쇼핑 행위 정보를 분석하여 <표 1>과 같이 고객 프로파일을 형성한다. <표 1>을 보면 알 수 있듯이 구매횟수에 기반한 기존 방법으로 고객 프로파일을 형성하면 (c)와 같이 데이터 희박성문제가 발생하는 것을 이 연구에서 제안하는 방법으로 형

<표 1> 웹마이닝과 상품계층도를 고려한 고객 프로파일 형성의 예

(a)  $d_{ij}^a$ : 고객  $i$ 의 상품  $j$ 에 대한 상품상세보기 횟수

CID	풋/핸드 케어	바디로션	로션	스킨	에센스	립스틱	파우더	향수류 남성용	향수류 여성용	향수세트
203	100	120	80	5	5	0	0	0	0	0
205	1	1	0	0	0	200	250	0	0	3
212	0	0	0	0	0	0	0	30	45	46
217	75	75	80	60	65	80	70	70	60	60
218	0	0	0	4	4	4	3	4	3	4

(b)  $d_{ij}^b$ : 고객  $i$ 의 상품  $j$ 에 대한 장바구니에 담은 횟수

CID	풋/핸드 케어	바디로션	로션	스킨	에센스	립스틱	파우더	향수류 남성용	향수류 여성용	향수세트
203	3	3	1	0	0	0	0	0	0	0
205	0	0	0	0	0	5	4	0	0	0
212	0	0	0	0	0	0	0	1	6	6
217	0	0	0	2	2	1	0	1	1	2
218	0	0	0	2	2	2	0	2	0	2

(c)  $d_{ij}^c$ : 고객  $i$ 의 상품  $j$ 에 대한 구매 횟수

CID	풋/핸드 케어	바디로션	로션	스킨	에센스	립스틱	파우더	향수류 남성용	향수류 여성용	향수세트
203	1	1	1	0	0	0	0	0	0	0
205	0	0	0	0	0	2	1	0	0	0
212	0	0	0	0	0	0	0	1	0	0
217	0	0	0	0	0	0	0	0	1	0
218	0	0	0	0	2	0	0	0	0	0

(d)  $d_{ij}$ : 식 (2)에 따라 형성된 고객 프로파일

CID	풋/핸드 케어	바디로션	로션	스킨	에센스	립스틱	파우더	향수류 남성용	향수류 여성용	향수세트
203	2.833	3	2	.042	.042	0	0	0	0	0
205	.004	.004	0	0	0	2.8	2.3	0	0	.012
212	0	0	0	0	0	0	0	1.82	1.98	2
217	.75	.75	1	1	1.25	1.5	.5	1	1.5	1
218	0	0	0	2	3	2	.75	2	.75	2



성된 프로파일 (d)와 비교해 보면 알 수 있다. 이 연구에서 사용하고 있는 방법은 Lee et al.[2001], Cho and Kim[2004] 등이 제안한 일반적인 인터넷 쇼핑몰에서의 상품구매 행동패턴은 상품상세보기, 장바구니 담기, 구매 등 3가지 연속적인 단계로 모형화 된다는 것에 근거하고 있다. 이러한 모형에 기반하여 고객의 선호도는 논리적으로 {클릭된 적이 없는 상품}  $\pi$  {클릭 되었던 상품}  $\pi$  {장바구니에 담겨 있던 상품}  $\pi$  {구매된 상품} 순으로 고객의 선호도가 높아질 것이다. 따라서 기존 연구에서 간과하였던, 상품상세보기 정보와 장바구니에 담은 정보 또한 해당 고객이 특정 상품들에 대한 선호도를 반영한다고 볼 수 있다. 따라서 고객선호도는 식 (2)와 같이 정의할 수 있다.

$P_{ij}^c$ 는 특정 기간동안 고객  $i$ 가 상품군  $j$ 에 속한 상품을 상품상세보기한 총횟수이고,  $P_{ij}^b, P_{ij}^p$ 도 고객  $i$ 가 상품군  $j$ 에 속한 상품을 장바구니에 담은 총횟수, 구매한 총횟수로 정의한다. 따라서 이 연구에서는 고객  $i$ 가  $j$ 번째 상품군에 대한 고객 프로파일 행렬  $P = (p_{ij}), i = 1, \dots, M$  ( $M$ 은 총 고객 수),  $j = 1, \dots, |G|$  ( $|G|$ 는 총 상품군 수)는 다음 식 (2)와 같이 정의된다.

$$\begin{aligned}
 P_{ij} &= \frac{p_{ij}^c - \min_{1 \leq j \leq |G|} (p_{ij}^c)}{\max_{1 \leq j \leq |G|} (p_{ij}^c) - \min_{1 \leq j \leq |G|} (p_{ij}^c)} \\
 &= \frac{p_{ij}^b - \min_{1 \leq j \leq |G|} (p_{ij}^b)}{\max_{1 \leq j \leq |G|} (p_{ij}^b) - \min_{1 \leq j \leq |G|} (p_{ij}^b)} \\
 &= \frac{p_{ij}^p - \min_{1 \leq j \leq |G|} (p_{ij}^p)}{\max_{1 \leq j \leq |G|} (p_{ij}^p) - \min_{1 \leq j \leq |G|} (p_{ij}^p)} \quad (2)
 \end{aligned}$$

식 (2)에서  $P_{ij}^c, P_{ij}^b$  그리고  $P_{ij}^p$ 를 그대로 사용하지 않고 각각 정규화(normalization)한 이유는  $P_{ij}^b$ 보다  $P_{ij}^p$ 가 그리고  $P_{ij}^c$ 가 일반적으로 매우 큰 값을 갖게 되므로 고객의 선호도가  $P_{ij}^c$ 에

의해 결정될 수 있기 때문이다. <표 1>은 웹마이닝과 상품계층도를 고려한 고객 프로파일 형성의 예를 보여주며, 식 (2)에 따라서 <표 1>(d)의  $P_{11}$  값은 다음과 같이 구해진다. 여기에 나타난 예는 계산편의상 고객 5명과 낮은 수준의 그래인을 임의로 선택하여 계산하였다.

$$P_{11} = \frac{100-0}{120-0} + \frac{3-0}{3-0} + \frac{1-0}{1-0} = 2.833$$

### 3.3 단계 3: 이웃 집단 탐색

이웃 집단 탐색 단계는 고객들간의 유사도를 계산하여 목표고객과 유사한 선호도를 지닌 고객들을 발견하는 단계로 단계 2와 함께 협업필터링 기반 상품추천방법에서 핵심과정이라고 할 수 있다. 먼저 고객들간의 유사도를 구하고 이를 바탕으로 이웃 집단들을 결정하게 된다. 대부분의 경우 두 고객  $a$ 와  $b$ 의 유사도를 측정하는 방법으로는 통계적 방법인 피어슨 상관계수를 이용하거나 정보검색분야에서 사용되는 코사인 함수를 이용하고 있다. 이웃을 형성하는 세부단계를 살펴보면 고객 프로파일  $P$ 가 주어졌을 때, 두 고객  $a$ 와  $b$ 의 유사도  $sim(a, b)$ 는 다음과 같은 두 가지 방법으로 측정될 수 있다.

- 상관계수: 피어슨 상관계수  $corr_{ab}$ 는 -1과 1사이의 값을 가지며 값이 1에 가까울수록 상관관계가 높다고 할 수 있다. 두 고객  $a$ 와  $b$ 의 상관계수는 다음 식 (3)을 이용하여 구한다.

$$\begin{aligned}
 sim(a, b) &= corr_{ab} \\
 &= \frac{\sum_j (p_{aj} - \bar{p}_a)(p_{bj} - \bar{p}_b)}{\sqrt{\sum_j (p_{aj} - \bar{p}_a)^2 \sum_j (p_{bj} - \bar{p}_b)^2}} \quad (3)
 \end{aligned}$$

여기서  $p_{aj}$ 와  $p_{bj}$ 는 상품  $j$ 에 관한 두 고객  $a$ 와  $b$ 의 평가점수이고,  $\bar{p}_a$ 와  $\bar{p}_b$ 는 두 고객  $a$ 와  $b$ 의 전체 상품의 평가점수 평균이다.

<표 2> CID203 고객의 유사도 계산의 예

CID	풋/핸드 케어	바디로션	로션	스킨	에센스	립스틱	파우더	향수류 남성용	향수류 여성용	향수세트	유사도 (CID203)
203	2.833	3	2	.042	.042	0	0	0	0	0	
205	.004	.004	0	0	0	2.8	2.3	0	0	.012	0.0014
212	0	0	0	0	0	0	0	1.82	1.98	2	0
217	.75	.75	1	1	1.25	1.5	.5	1	1.5	1	0.4173
218	0	0	0	2	3	2	.75	2	.75	2	0.0089

- 코사인: 두 고객  $a$ 와  $b$ 의 유사도는 두 벡터 사이의 코사인 값을 이용하여 계산할 수 있다. 코사인  $\cos(\vec{a}, \vec{b})$ 은 다음 식 (4)와 같이 구한다.

$$\begin{aligned}
 sim(a, b) &= \cos(\vec{a}, \vec{b}) = \frac{\mathbf{P}_a \cdot \mathbf{P}_b}{\|\mathbf{P}_a\| \|\mathbf{P}_b\|} \\
 &= \frac{\sum_{k=1}^{|G|} p_{ak} \times p_{bk}}{\sqrt{\sum_{k=1}^{|G|} p_{ak}^2} \times \sqrt{\sum_{k=1}^{|G|} p_{bk}^2}}
 \end{aligned}
 \tag{4}$$

여기서  $\mathbf{P}_a$ 와  $\mathbf{P}_b$ 는 두 고객  $a$ 와  $b$ 의 고객 프로파일의 행벡터(row vector)이다. 현재까지 두 방법간에는 큰 차이가 없는 것으로 알려져 있다 [Sarwar et al., 2000b]. 따라서 이 연구에서는 고객간의 유사도 측정을 위하여 코사인 함수를 사용하기로 한다. <표 2>는 전 단계에서 형성된 고객 프로파일을 기반으로 CID203 고객과 다른 고객과의 유사도를 계산하는 예를 보여주고 있다. 제시된 프로파일만 고려할 때, CID203 고객과 가장 유사한 고객은 CID217 고객임을 알 수 있다.

### 3.4 단계 4: 추천 상품 결정

추천 상품 결정 단계에서는 목표고객에게 추천할 상품을 결정한다. 유사한 선호도를 가진 이웃들이 선호하는 상품 중 목표고객이 기존에 구

매한 상품을 제외하고 가장 구매할 가능성이 높은 상위 N개의 상품추천 목록을 생성한다. 목표 고객을 위한 추천상품을 판단하는 기준으로 다음 3가지 방법을 일반적으로 사용한다.

- 빈발 구매 상품(MFP, Most frequently purchased product): Sarwar et al.[2000]의 연구에서 판매 데이터베이스에서 이웃들의 구매 빈도가 높은 상품을 추천하는 방법으로, 구매 빈도가 높은 순으로 정렬하여 상위 N개의 상품 중 목표고객이 구매하지 않은 상품을 추천한다. 이 방법은 인기있는 상품일수록 구매 가능성이 더 높다는 경험론적인 지식에 근거한 방법이다.
- 빈발 참조 상품(MFR, Most frequently referred product): 참조빈도(reference frequencies)에 따라 상품을 분류하여, 참조빈도가 높은 순으로 상품을 추천한다. 특정 상품  $j$  ( $1 \leq j \leq N$ )에 관한 고객  $a$ 의 참조빈도  $RF_{a,j}$ 는 다음과 같이 정의할 수 있다.

$$\begin{aligned}
 RF_{a,j} &= \sum_{i \in \text{고객 } a \text{의 이웃}} \frac{r_{ij}^c - \min_{1 \leq j \leq N}(r_{ij}^c)}{\max_{1 \leq j \leq N}(r_{ij}^c) - \min_{1 \leq j \leq N}(r_{ij}^c)} \\
 &+ \frac{r_{ij}^b - \min_{1 \leq j \leq N}(r_{ij}^b)}{\max_{1 \leq j \leq N}(r_{ij}^b) - \min_{1 \leq j \leq N}(r_{ij}^b)} \\
 &+ \frac{r_{ij}^p - \min_{1 \leq j \leq N}(r_{ij}^p)}{\max_{1 \leq j \leq N}(r_{ij}^p) - \min_{1 \leq j \leq N}(r_{ij}^p)}
 \end{aligned}
 \tag{5}$$

여기에서  $N$ 은 총 상품 수이고,  $r_{ij}^c, r_{ij}^b, r_{ij}^d$ 는 각각 고객  $i$ 가 상품  $j$ 를 상품상세보기 단계, 장바구니 담기 단계, 구매 단계에서의 발생빈도를 나타낸다. 이 방법은 많이 참조된 상품일수록 고객이 선호할 것이라는 가정하에 고객 프로파일을 생성할 때 사용한 클릭스트림 데이터를 이용하여 계산한다.

- 구매전환비율이 높은 상품 추천(HCR, Highest click-to-buy rate): 이 방법은 구매전환비율이 높은 상품이 마케팅 효과가 더 크다는 가정하에[Lee et al., 2001] 모든 이웃의 상품상세보기 대 구매전환비율(click-to-buy)이 높은 상품을 추천한다. 다시 말하면 전환비율(click-to-buy)은 상품상세보기를 한 후 결과적으로 그 상품이 구매로 전환된 비율을 의미한다. 상품  $j(1 \leq j \leq N)$ 를 상품상세보기 한 후 구매로 전환한 고객  $a$ 의 구매전환비율  $CTB_{a,j}$ 는 다음과 같이 정의할 수 있다.

$$CTB_{a,j} = \frac{\sum_{i \in \text{고객 } a \text{의 이웃}} r_{ij}^d}{\sum_{i \in \text{고객 } a \text{의 이웃}} r_{ij}^c} \quad (6)$$

여기에서  $N$ 은 총 상품 수이고,  $r_{ij}^c, r_{ij}^b$ 는  $i$  고객이 상품  $j$ 를 상품상세보기 단계, 구매 단계에서의 발생빈도를 말한다.

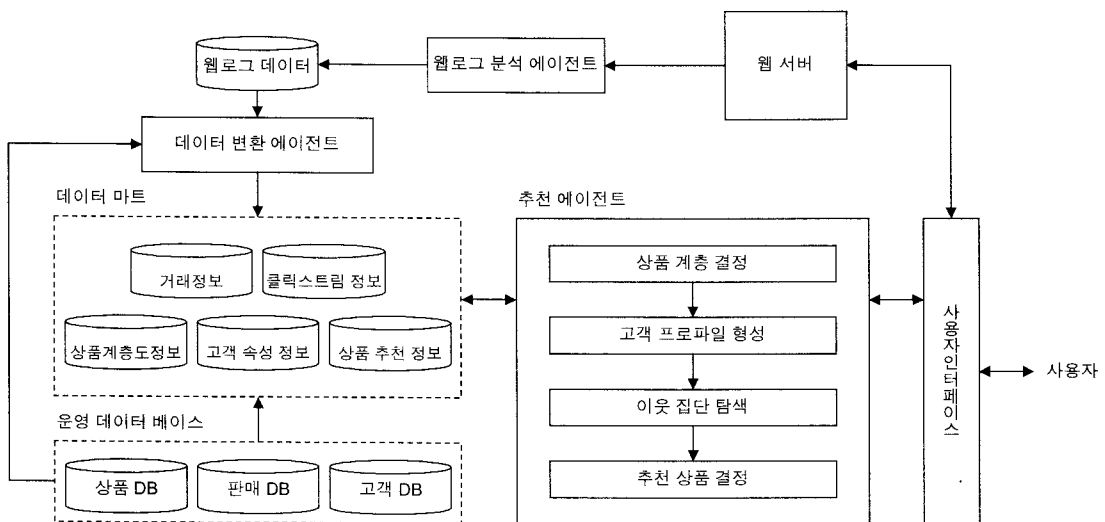
어느 방법을 사용하는가는 분야별로 결과가 다르기 때문에 주로 휴리스틱한 방법으로 실험을 통하여 결정한다. 이 연구에서는 Cho and Kim[2004]이 제안한 MFR 방법을 이용한다.

## IV. WebCF-PT 추천시스템

### 4.1 시스템 구조도

이 장에서는 III장에서 제시한 상품추천방법론을 기반으로 인터넷 쇼핑물을 위한 WebCF-PT 추천시스템의 구조도 및 각 구성 에이전트들에 대하여 설명하고자 한다. <그림 4>는 WebCF-PT 추천시스템의 구조도이다. 그림에서 볼 수 있듯이 웹로그 분석 에이전트(Web Log Analysis Agent), 데이터 변환 에이전트(Data Transformation Agent), 추천 에이전트(Recommender Agent)로 구성되어 있다.

각 구성요소들의 기능을 간단히 살펴보면, 먼



<그림 4> WebCF-PT 추천시스템 구조도

저 웹로그 분석 에이전트가 방대한 양의 웹로그 파일을 전처리하여 웹로그 데이터베이스에 저장하면, 데이터 변환 에이전트는 운영 데이터베이스와 웹로그 데이터베이스로부터 상품추천에 사용될 데이터를 추출, 정제, 변환하여 데이터 마트에 저장한다. 추천 에이전트는 추천 알고리즘을 통해 고객별로 개인화된 상품추천 목록을 작성하고 고객 선호도, 추천에 사용된 데이터, 그리고 실시간 추천을 제공하기 위한 추천목록을 데이터 마트에 저장한다. 각 구성요소에 대한 자세한 설명은 다음 4.2절에서 소개한다.

## 4.2 시스템 구성요소

WebCF-PT 추천시스템을 구성하고 있는 각 에이전트들의 역할에 대해 자세히 살펴보면 아래와 같다.

- 웹로그 분석 에이전트: 웹 서버에 있는 텍스트 형태의 웹로그 파일은 곧바로 유용한 정보가 되지 못하므로 전처리를 해야하며, 그 결과 산출된 데이터를 데이터베이스에 저장하게 된다. 데이터를 적재하기 전의 전처리 과정은 데이터정제(Data Cleaning), 사용자확인(User Identification), 세션식별(Session Identification), 클릭스트림 경로완성, 포매팅(Formatting), 데이터 로딩(Data Loading)으로 구성된다. 웹로그 분석 방법은 Cooley et al. [1999]이 제안한 방식을 채택한다. 웹로그 분석 에이전트의 전처리 단계를 자세히 살펴보면 다음과 같다.
  - 데이터정제: html파일에는 그림, 영상, 스크립트와 같은 불필요한 아이템들이 많이 포함되어 있다. 확장자가 jpg, gif, cgi등으로 끝나는 파일은 삭제하고 인터넷 쇼핑몰의 쇼핑단계에 맞는 URL을 매핑하여 필요한 파일만을 추출한다.
  - 사용자확인: 일반적으로 사용자확인 기준으로

로 동일한 IP가 이용되나 이 연구에서는 고객아이디를 기준으로 사용자를 확인한다.

- 세션식별: 세션(session)이란 처음 사이트에 진입하여 최종적으로 사이트를 빠져나가는 과정으로 정의된다. 그렇지만 방문 고객이 현재 웹 페이지가 아닌 다른 페이지에서 작업 또는 다른 브라우저를 실행하는 경우 실제로는 그 인터넷 쇼핑몰을 사용하고 있지 않으나 로그 파일은 해당 페이지를 보고 있는 것으로 간주한다. 이는 로그 파일에 시작과 종료가 명확히 구분 되지 않기 때문이다. 따라서 페이지 요청(page request)이 30분 이상 경과하면 세션이 종료됐다고 간주한다. 그리고 로그파일의 세션은 유일함으로, 이 연구에서는 하나의 세션을 고객의 쇼핑 행위의 기본단위(transaction)로 구분한다.
- 클릭스트림 경로 완성: 웹 사이트는 다양한 페이지의 링크로 구성 되어 있으며 하나의 페이지는 다른 여러 페이지와 연결되어 있다. 따라서 사이트의 구성을 분석하게 되면 사용자가 어떤 경로로 사이트를 브라우징 하는지 파악할 수 있다. 이 연구에서는 로그의 쇼핑 경로를 잃어버렸을 때 웹로그상의 참조 페이지(referrer log)를 이용하여 잃어버린 페이지를 유추하여 경로를 완성한다.
- 포매팅: 웹로그를 파싱하여 분석에 적합한 포맷으로 변경한다.
- 데이터 로딩: 전 처리한 데이터를 운영데이터베이스처럼 접근 및 분석 가능하도록 웹로그 데이터베이스에 저장한다.
- 데이터 변환 에이전트: WebCF-PT 추천시스템의 데이터 마트를 유지 관리하기 위해 운영 데이터베이스와 웹로그 데이터베이스로부터 자료를 추출, 정제하고 추천시스템에 적용하기 적합한 형태로 데이터를 변환시키는 과정을 수행한다. 데이터 변환 에이전트는 운영 데이터베이스의 고객 정보로부터 분석에 필요한 데이터를 추출하여 데이터 마트의 고객 속성

정보에 저장하고 운영 데이터베이스의 판매 데이터와 상품 데이터로부터 상품 정보, 상품 클래스 정보, 구매 빈도를 추출하여 거래 정보에 저장한다. 그리고 웹로그 데이터베이스로부터 고객별 쇼핑 단계 정보와 상품코드를 추출하여 클릭스트림 데이터에 저장한다. 이때 상품 코드가 없는 로그정보는 삭제 하고 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.

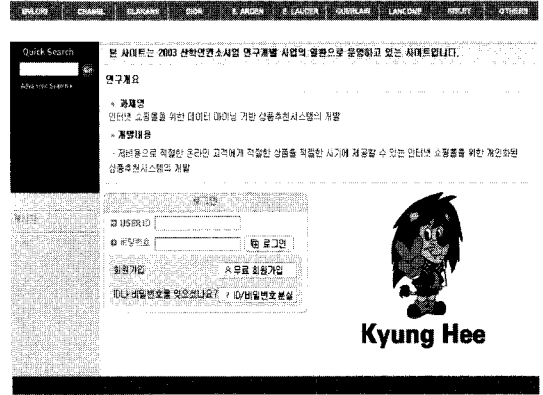
- 추천 에이전트: 추천 에이전트는 데이터 마트의 상품 관련 정보와 고객 관련 정보를 이용하여 구현된 개인화된 추천목록을 생성한다. 그리고 상품추천과정에서 발생한 고객 프로필 정보, 이웃고객 정보, 고객 선호도 정보, 추천 상품정보를 데이터 마트에 저장한다.

### 4.3 시스템 구현

WebCF-PT추천시스템을 프로토타입 시스템으로 개발하고 이를 적용하기 위하여 실험목적의 EBIB Research 쇼핑몰을 개발하였다. <그림 5>는 쇼핑몰 초기화면을 보여주고 있다. WebCF-PT추천시스템 프로토타입의 데이터 베이스시스템으로 MS-Access를 사용하였고, 각 에이전트와 구성요소들은 Visual Basic6.0을 사용하여 개발하였다. 그리고 가상 인터넷 쇼핑몰인 EBIB Research는 데이터 베이스시스템으로 MS- SQL Server 2000을 사용했고, WindowsNT2000 기반의 IIS5.0 웹 서버와 ASP(Active Server Page)를 이용하여 개발하였다.

웹로그 데이터를 얻기 위하여 MS IIS가 제공하는 여러 로그파일 중 W3C Extended 로그파일을 사용하였고, 웹 서버부터 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.



<그림 5> EBIB Research 쇼핑몰 초기화면

EBIB Research 쇼핑몰에는 287개의 상품이 소개되어 있다. 상품계층은 루트 노드를 제외한 3개의 계층으로 구성되어 있다. 상위 계층은 7개의 상품클래스가 있고 중간 계층은 37개의 상품클래스로 구성되어 있고, 하위 계층은 287개의 개별상품으로 구성되어 있으며, 또한 고객에게 적시에 상품추천을 제공할 수 있도록 목표고객별 이웃의 고객ID를 저장하고 있도록 하였다. 그리고, 개인별 상품추천 목록을 제공받을 고객 ID와 상품코드를 데이터 마트에 저장하였다.

## V. WebCF-PT에 의한 상품 추천 사례

WebCF-PT추천시스템의 이해를 돕기 위하여 한 고객이 상품을 구매하고 추천을 받는 과정을 설명하고자 한다. 고객 Kim은 회원가입을 하기 위해 <그림 5>에 나타난 EBIB Research 쇼핑몰 초기화면에서 사용자 ID, 이름, e-mail, 생년월일, 성별, 주소와 같은 데이터를 입력한다. Kim에게는 사이버 머니 백 만원이 지급되고, 사이버 머니

한도 내에서 원하는 상품을 구매 할 수 있도록 한다. 회원가입이 성공적으로 끝나면 자동적으로 로그인되어 EBIB Research 쇼핑몰에 접속한다. 다음에 방문할 때는 회원가입 없이 개인화된 서비스를 받을 수 있다. <그림 6>은 EBIB Research 쇼핑몰의 메인 화면이다.

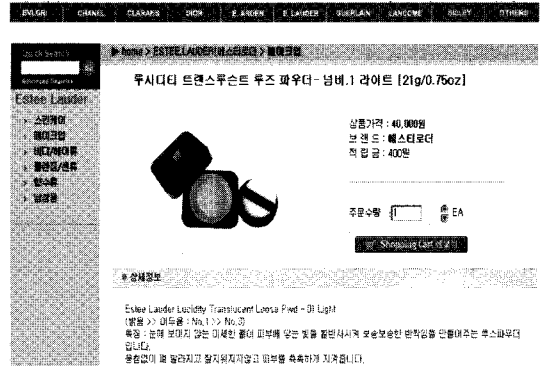


<그림 6> 메인 화면

고객 Kim이 회원가입을 한 후 웹사이트를 방문하고 상품구매를 하기위해 웹서핑을 하면, 해당 웹서버는 자동적으로 회원의 방문과 관련된 로그가 저장된다. WebCF-PT추천시스템은 웹로그 에이전트를 이용하여 웹로그 파일을 전처리 한 후, 상품상세보기는 'goods\_view.asp', 장바구니담기는 'cart\_list.asp', 구매는 'cart\_pay.asp'로 매핑하고, 쇼핑행위와 관련된 데이터를 추출한다.

<그림 7>에서 Kim은 스킨케어 상품군이 2.57로 가장 높은 선호도 점수를 얻었고, 기타 상품군이 2.41로 두 번째로 높은 선호도를 보였다. Kim의 프로파일은 III장에서 설명한 혼합수준 그래픽인 <그림 3>(c)과 웹마이닝을 기반으로 구하였으며, 이웃고객은 Kim의 프로파일을 사용하여 구하였다. Cho and Kim(2004)은 이웃의 수는 50명 정도 일 때 좋은 성능을 보이고, 추천방법 중 MFR이 좋은 성과를 보인다는 것을 실험을 통해 증명하였다. 본 연구에서는 가입자 수가 충분하지 않으므로 50명일 때와 큰 차이가 없는 30명을 이

웃의 수로 선택 하였고, MFR추천방법을 사용하였다. 상품 참조빈도가 높은 상품 중 Kim이 과거에 구매한 상품을 제외한 상위 10개를 선택하여 추천한다. <그림 8>은 Kim에게 제공된 개인화된 추천상품리스트를 보여주고 있다.



<그림 7> 상품상세보기 화면

1		상품명 가격 제조사	ring 280,000 샤넬
2		상품명 가격 제조사	알마니 시계 AR-0224 219,000 알마니
3		상품명 가격 제조사	벌벳 소프트 클렌징 밀크 [200ml/6.7oz] 33,800 불가리
4		상품명 가격 제조사	뎀 데오그란트 스킨 [75ml/2.7oz] 21,900 크리스찬디올
5		상품명 가격 제조사	플러저 바다 로션 [250ml/8.3oz] 46,900 에스티로더
6		상품명 가격 제조사	뷰티풀 바다 로션 [250ml/8.3oz] 47,000 에스티로더
7		상품명 가격 제조사	보태니컬 썬 블록 spf 20 —무색 40ml/1.4oz [40ml/1.4oz] 115,600 시슬리
8		상품명 가격 제조사	샤넬 샹스 오 드 투와렛 피플 73,800 샬랑
9		상품명 가격 제조사	포도제닉 라이트 리플렉팅 겔 매이컵 100 30ml/1oz [30ml/1oz] 36,900 랑콤
10		상품명 가격 제조사	데피니션-넘베.1 눈와르 앵피 니 [6.5ml/0.21oz] 28,100 랑콤

<그림 8> 고객 Kim에게 제공된 상품추천리스트

## VI. 결 론

전자상거래의 급성장으로 기업들의 생존을 위한 경쟁은 더욱 심화되어 다른 경쟁업체보다 경쟁우위를 가질 수 있는 마케팅 전략이 필요하게 되었고, 고객은 상품 정보의 과다로 인하여 효과적으로 상품을 선택할 수 없게 되는 상품 과부하 현상을 야기시켰다. 이러한 문제를 해결하기 위한 정보 기술 중의 하나가 고객의 선호도에 부합하는 상품을 찾도록 도와주는 상품추천시스템이다. 현재까지 상품추천시스템을 구현하기 위한 다양한 기술들이 개발되어 왔는데, 이 중에서 협업필터링이 가장 성공적인 상품추천 기술로 알려져 있다. 본 연구에서는 협업필터링의 데이터 희박성문제와 시스템 확장성문제들을 해결하기 위하여 웹마이닝과 상품계층도를 이용한 협업필터링 기반 개인별 상품추천시스템, WebCF-PT 프로토타입 시스템을 개발하고 이를 적용하기 위하여 실험 목적의 EBIB Research 쇼핑몰을 구축하였다. 이 연구에서 구축한 쇼핑몰에서 고객이 상품 둘러보기를 하고 클릭하고 장바구니에 담고, 사이버머니로 상품을 구매하게 되면 그러

한 고객정보를 바탕으로 제공한 추천 상품리스트와 고객의 구매 데이터만을 이용한 추천 상품리스트를 고객들과의 인터뷰를 통하여 비교하였다. 다양한 정보를 바탕으로 제시한 상품추천리스트가 더욱 고객의 선호도가 부합되는 것으로 결과가 나타났지만, 피실험자 수가 충분하지 않았고 비교 분석은 이번 연구의 범위에서 벗어나므로 추후 연구과제로 하기로 한다. 비교분석 방법으로는 고객들로부터 구매하고 싶은 목록을 제공 받고 이를 WebCF-PT에 의한 추천 목록과 비교해보면 WebCF-PT시스템의 효과성을 검증할 수 있는 좋은 방법으로 생각되며, 추후 연구에서 EBIB Research 쇼핑몰을 더욱 보강하여 실험을 할 예정이다. 또한 고객에게 상품을 추천하였을 때, 고객이 추천한 이유를 설명해 줄 수 있으면 더욱 추천의 효과를 높일 것으로 기대되며 이에 대한 연구가 필요하며, 기존의 CF시스템과 실험을 통한 성능을 비교 및 분석할 필요가 있다. 또한 인터넷 쇼핑몰뿐만 아니라 다양한 영역에서 추천방법간의 차이가 어떻게 달라지는지를 비교 분석 하는 것도 좋은 연구가 될 것으로 기대된다.

## 〈참 고 문 헌〉

- [1] 김재경, 안도현, 조윤희, "Development of a Personalized Recommendation Procedure Based on Data Mining Techniques for Internet Shopping Malls," *한국지능정보시스템학회*, 제9권 제3호, 2003, pp. 177-191.
- [2] 김재경, 서지혜, 안도현, 조윤희, "A Personalized Recommendation Methodology based on Collaborative Filtering," *한국지능정보시스템학회*, 제8권 제2호, 2002, pp. 139-157.
- [3] Adomavicius, G. and Tuzhilin, A., "Expert-driven Validation of Rule-Based User Models in Personalization Applications," *Data Mining and Knowledge Discovery*, Vol. 5, 2001, pp. 33-58.
- [4] Basu, C., Hirsh, H. and Cohen, W., "Recommendation as classification: Using social and content-based information in recommendation," In Proc. *15th National Conference on Artificial Intelligence*, 1998, pp. 714-720.
- [5] Berry, J.A. and Linoff, G., *Data mining techniques: for marketing, sales, and customer support*, New York: Wiley, 1997.
- [6] Billsus, D. and Pazzani, M.J., "Learning

- collaborative information filters," In Proc. *15th International Conference on Machine Learning*, 1998, pp. 46-45.
- [6] Cho, Y.H. and Kim, J.K., "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, Vol. 26, 2004, pp. 233-246.
- [7] Cooley, R., Mobasher, B. and Srivastava, J., "Data Preparation for Mining World Wide Web Browsing Patterns," *Knowledge and Information Systems*, Vol. 1, No. 1, 1999.
- [8] Han, J. and Fu, Y., "Mining Multiple-Level Association Rules in Large Databases," *IEEE Transaction on Knowledge and Data Engineering*, Vol. 11, No. 5, 1999, pp. 798-804.
- [9] Herlocker, J.L., Konstan, J.A. and Riedl, J., "Explaining Collaborative Filtering Recommendations," In Proc. *2000 ACM Conference on Computer Supported Cooperative Work*, 2000, pp. 241-250.
- [10] Herlocker, J.L., Konstan, J.A., Borchers, A. and Riedl, J., "An Algorithmic Framework for Performing Collaborative Filtering," In Proc. *Conference on Research and Development in Information Retrieval*, 1999, pp. 230-237.
- [11] Hill, W., Stead, L., Rosenstein, M. and Furnas, G.W., "Recommending and Evaluating Choices in a Virtual Community of Use," In Proc. *Human Factors in Computing Systems*, 1995, pp. 194-201.
- [12] Kim, J.K., Cho, Y.H., Kim, W.J., Kim, J.R. and Suh, J.Y., "A Personalized Recommendation Procedure for Internet Shopping Support," *Electronic Commerce Research and Applications*, Vol. 1, No. 3, 2002, pp. 301-313.
- [13] Lee, J., Podlaseck, M., Schonberg, E. and Hoch, R., "Visualization and analysis of Clickstream data of online stores for understanding Web merchandising," *Data Mining and Knowledge Discovery*, Vol. 5, No. 1, 2001, pp. 59-84.
- [14] Lawrence, R.D., Almasi, G.S., Kotlyar, V., Viveros, M.S. and Duri, S.S., "Personalization of supermarket product recommendations," *Data Mining and Knowledge Discovery*, Vol. 5, No. 1, 2001, pp. 11-32.
- [15] Melville, P., Mooney, R.J. and Nagarajan, R., "Content-boosted Collaborative Filtering," In Proc. *SIGIR 2001 Workshop on Recommender Systems*, 2001.
- [16] Mobasher, B., Cooley, R. and Srivastava, J., "Automatic Personalization based on Web usage mining," *Communications of the ACM*, Vol. 43, No. 8, 2000, pp. 142-151.
- [17] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," In Proc. *Computer Supported Cooperative Work*, Chapel Hill, NC, 1994, pp. 175-186.
- [18] Sarwar, B., *Sparsity, scalability, and distribution in recommender systems*, Ph.D. thesis, University of Minnesota, 2001.
- [19] Sarwar, B., Karypis, G., Konstan, J.A. and Riedl, J., "Application of dimensionality reduction in recommender system - a case study," In Proc. *ACM WebKDD-2000 Workshop*, 2000a.
- [20] Sarwar, B., Karypis, G., Konstan, J.A. and Riedl, J., "Analysis of recommendation algorithms for e-commerce," In Proc. *ACM E-Commerce*, 2000b, pp. 158-167.



## ◆ 저자소개 ◆



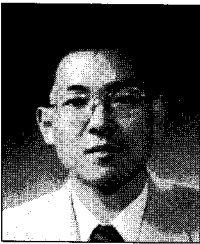
김재경 (Kim, Jae Kyeong)

현재 경희대학교 경영대학 e비즈니스전공 부교수 및 경영대학원 e비즈니스 전공주임교수로 재직중이다. 서울대학교 산업공학과를 졸업하고, 한국과학기술원에서 경영정보시스템 전공으로 석사 및 박사학위를 취득하였으며, 미국 미네소타 주립대학교 경영정보학과에서 교환교수, 경기대학교 경영정보학과에서 조교수를 역임하였다. 연구분야로는 e-CRM, e비즈니스 모형 및 전략, 의사결정분석, 상품추천시스템 등이다.



안도현 (Ahn, Do Hyun)

현재 경희대학교 사회과학연구원 정보센터에서 책임연구원으로 재직 중이다. 경희대학교 사회학과에서 학사, 경영학과에서 MIS 전공으로 석사학위를 취득하고 동 대학원 경영학과 e-Business 전공 박사과정을 수료하였다. 주요 관심분야는 Data Mining, e-Commerce, 상품추천시스템, m-Commerce 등이며, International Journal of Internet and Enterprise Management, Lecture Notes in Artificial Intelligence, 한국경영과학회지, 한국지능정보시스템학회논문지 등에 논문을 게재하였다.



조윤호 (Cho, Yoon Ho)

현재 국민대학교 경상대학 e-비즈니스학부 조교수로 재직 중이다. 서울대학교 계산통계학과(전산학전공)를 졸업하고, KAIST 경영정보공학과에서 석사학위를 취득하였으며 KAIST 경영공학과에서 박사학위를 취득하였다. LG전자(주)에서 6년간 주임연구원으로 재직하였다. 주 연구 분야는 e-Business, Mobile business, CRM, Data mining 등이며, IEEE Intelligent Systems, Expert Systems with Applications, Electronic Commerce Research and Applications, Computers and Industrial Engineering, International Journal of Internet and Enterprise Management 등에 논문을 게재하였다.

◆ 이 논문은 2004년 4월 28일 접수하여 1차 수정을 거쳐 2004년 11월 1일 게재확정되었습니다.