

지수적 가중치를 적용한 협력적 상품추천시스템

이 경 희[†] · 한 정 혜^{††} · 임 춘 성^{†††}

요 약

전자상점에서 이루어지는 고객의 구매패턴이 온라인 상에서 데이터베이스화되어, 이를 통하여 고객의 취향에 맞는 상품을 제공할 수 있는 많은 알고리즘이 연구되고 있다. 이러한 알고리즘은 전자상점에서 고객의 개별특성을 고려한 상품을 제공하기 위하여, 고객정보 데이터베이스와 거래정보 데이터베이스로부터 연관규칙 등을 추출하여 사용한다. 그러나 시간의 흐름에 민감한 계절상품이나 특선상품과 같이 전자상점의 거래량에 크게 직결될 수 있는 상품에도 기존의 시간을 고려하지 않은 알고리즘을 적용한다면 추천성공률이 떨어질 것이다. 따라서 본 논문에서는 시간의 영향을 많이 받는 상품추천을 위하여, 최근 전자상점 추천시스템으로 효과적인 아이템 기반 협력알고리즘에 지수적 가중치를 적용한 협력적 여과추천(EWCFR) 알고리즘을 제안한다. 또한 이러한 추천시스템이 대용량의 고객데이터와 상품데이터에 대한 연산을 수행하고 다수의 고객에게 실시간으로 서비스를 제공하여야 하므로, XML기반의 MMDB를 활용한 전자상거래 시스템과 알고리즘을 제안한다.

A Recommendation System of Exponentially Weighted Collaborative Filtering for Products in Electronic Commerce

Kyung Hee Lee[†] · Jeong Hye Han^{††} · Choon Seong Leem^{†††}

ABSTRACT

The electronic stores have realized that they need to understand their customers and to quickly response their wants and needs. To be successful in increasingly competitive Internet marketplace, recommender systems are adapting data mining techniques. One of most successful recommender technologies is collaborative filtering (CF) algorithm which recommends products to a target customer based on the information of other customers and employ statistical techniques to find a set of customers known as neighbors. However, the application of the systems, however, is not very suitable for seasonal products which are sensitive to time or season such as refrigerator or seasonal clothes. In this paper, we propose a new adjusted item-based recommendation generation algorithms called the exponentially weighted collaborative filtering recommendation (EWCFR) one that computes item-item similarities regarding seasonal products. Finally, we suggest the recommendation system with relatively high quality computing time on main memory database (MMDB) in XML since the collaborative filtering systems are needed that can quickly produce high quality recommendations with very large-scale problems.

키워드 : 연관규칙(Association Rule), 아이템 기반 협력여과 추천알고리즘(Item-based Collaborative Filtering Recommendation Algorithm), 지수적 가중치(Exponential Weight), XML, MMDB

1. 서 론

현재 전세제는 최첨단 정보통신 기술의 발전으로 산업화 시대에서 고도의 정보화 사회로 급속한 도래기를 맞고 있으며, 값싸고 사용하기 쉬워 빠르게 전파되고 있는 인터넷의 등장으로써 전자적인 단일 상권으로 묶이는 본격적인 글로벌 전자상거래 시대로 돌입하고 있다. 실제로 국내외에

서 이미 수많은 전자상점이 활발히 운영 중이며, 무한경쟁의 상거래 속에서 살아남기 위하여 원투원 마케팅 개념을 적용한 CRM(Customer Relationship Management)을 도입하여 서비스를 제공하고 있다

고객웨어전략의 가장 중요한 원칙이며 핵심은 파레토 원리(Pareto Principle)-어떤 회사 비즈니스의 80%는 20%의 고객들로부터 창출된다-이다[1]. 실제 이동통신 등 많은 회사에서 CRM을 궁극적으로 이러한 로얄고객을 확보 및 유지하는 것을 목표로 많은 비용을 들이고 있다. 이렇게 고객군을 분류하여 로얄고객을 관리하는 고객기반에서 상품기

† 준 회 원 : 충북대학교 대학원 전자계산학과
 †† 종 신 회 원 : 청주교육대학교 컴퓨터교육학과 교수
 ††† 정 회 원 : 연세대학교 컴퓨터산업시스템공학과 교수
 논문접수 : 2001년 10월 8일, 심사완료 : 2001년 12월 7일

반으로 생각해보면, 인기상품의 매출이 전체 매출에 크게 영향을 주는 등 비슷하게 볼 수 있다. 또한 매출에 영향을 주는 상품의 다수는 유행과 시간에 따라 매우 민감하게 반응하므로, 이를 고려하여 CRM을 하게 된다면 효과적일 것이다. 그러나 아직까지 대부분 전자상점들의 추천시스템은 다수의 고객과 거래 데이터베이스의 분석을 통한 모든 고객과 상품을 대상으로 고정된 시점(time)하에서 연관규칙, 클러스터링 또는 룰을 계산하는 것이 거의 대부분이다[8].

따라서 본 논문은 전자상점의 매출에 영향이 크게 미치는 계절상품 또는 특별상품에 대한 추천시스템을 제안하고자 한다. 예를 들어, 계절상품 에어컨을 추천함에 있어서 에어컨의 가격대나 성능 및 용량 등의 특성 등 모델에 따라 구매하는 고객군이 다를 것이며, 이를 정확히 추천한다면 임의의 고객이 해당 계절상품을 구입할 의사는 더욱 커지게 되는 것이다. 이러한 추천 알고리즘으로는 고객들의 협력적 여과 추천(Collaborative Filtering Recommendation ; CFR) 알고리즘이 효과적으로 나타나는데[5], 즉 현재 로그인한 활동 고객과 가장 유사한 이웃고객들의 거래 데이터베이스를 토대로 패턴을 계산하여, 결과에 따라 인기상품 또는 특별상품 또는 계절상품을 차별화 제시할 수 있으므로 효과적인 CRM이 이루어지게 되는 것이다. 이때 계절상품은 시간에 매우 민감하므로, 최근의 이웃고객들의 거래데이터에 대한 가중치가 높고 예전 이웃고객들의 거래데이터에 대한 가중치를 낮게 적용하는 것이 효과적이다. 따라서 이를 위하여 [11]의 지수적 가중치를 적용한 협력적 여과(Exponentially Weighted Collaborative Filtering Recommendation ; EWCFR) 알고리즘을 제안하고자 한다.

그러나 이러한 상품들에 대한 고객의 반응(구매)은 시간에 매우 민감하게 변화하므로 고객거래 데이터베이스의 실시간 분석에 의한 추천을 해야 효과적이데, 실시간 분석처리는 많은 고객들이 이용하는 대규모 전자상점에 있어서 서버에 큰 부하를 주므로, 아직은 효과적으로 적용되지 못하고 있는 실정이다. 즉 개별추천을 위해 요구되어지는 상품과 고객간의 상호관련 규칙은 많은 데이터베이스 접근 및 계산이 요구되므로, 이러한 단점을 극복하기 위하여 서버의 데이터베이스에 물리적인 접근비용을 낮출 수 있는 최근에 각광받는 메인 메모리 데이터베이스(Main Memory Database ; MMDB)를 이용하고자 한다[9, 12].

이와 같이 본 논문에서는 고객 데이터베이스와 거래 데이터베이스를 분석하여 유사고객군으로 분류한 후, 해당 유사고객군의 고객데이터들을 EWCFR 알고리즘을 통하여 특정고객이 구매할 가능성이 가장 크면서 시간에 민감한 계절상품 또는 인기상품을 추천하는 시스템을 제안하고자 한다. 또한 이 계절상품 추천 시스템은 전자상점에 사용되는

모든 콘텐츠와 데이터를 시스템과 벤더에 독립적 개방표준인 XML로 표현하고, 개별화 서비스를 위한 빈번한 질의처리를 위해 MMDB를 활용하고자 한다.

논문의 구성은 먼저 2장에서 관련 선행연구를 살펴보고, 3장에서 시간에 민감한 계절상품을 추천하기 위한 전자상점의 구조와 지수적 가중치 협력적 알고리즘, 4장에서는 MMDB를 활용한 XML기반 추천시스템, 마지막으로 5장에서는 결론과 향후 연구를 제시하겠다.

2. 관련 연구

2.1 전자상점 추천 알고리즘

현재 전자상점에 적용할 수 있는 상품 소개법으로는 1990년대 초반 미국의 돈 페퍼스와 마사 로저스에 의해 처음으로 개념이 확립된 일대일 마케팅으로서, 관련상품의 일괄구매를 유도하는 크로스셀링, 동종이급 상품을 제안하는 업셀링, 그리고 타겟 셀링 등으로 나뉠 수 있다[1]. 실제로 Amazon과 같은 전자상점은 고객의 구매패턴과 고객 정보를 기반으로 프로파일을 구축한 뒤, 고객이 웹 사이트에 접속할 때 적절한 서적을 추천해 주는 에이전트 기법을 적용한 서비스를 제공하고 있다.

[3]은 전자상점에 적합한 컴포넌트 콘텐츠 레이아웃 설계를 위한 연구이며, 고객이 전자상점에 로그인 한 후 추천시스템에 의해 처리된 상품 검색결과는 다음 <표 1>와 같은 컴포넌트 콘텐츠 레이아웃으로 구성될 수 있다. 본 연구에서는 기획셀링의 상품을 추천하기 위한 알고리즘과 XML기반 MMDB 구현 설계방법을 제안하고자 한다.

<표 1> 개별화 레이아웃 컴포넌트 콘텐츠

레이아웃 구성요소	기능
기획셀링	시간의 변화에 민감한 계절상품, 특별기획상품을 추천
타겟 셀링	유사고객군의 협력적 상품거래 내역을 여과한 상품 중 최근접 상품 제공
동시 크로스셀링	유사고객군의 장바구니 연관규칙을 협력적 여과한 상품 중 최근접 상품 제공
순차 크로스셀링	유사고객군의 협력적 기존 구매상품 거래중 다른 상품디렉토리를 여과한 상품 중 최근접 상품 제공
업셀링	유사고객군의 기존 구매상품의 같은 상품 디렉토리내 상위거래를 협력적 여과한 상품 중 최근접 상품 제공

이러한 고객의 신상명세와 구매행동을 바탕으로 개별화된 셀링 기법을 통한 고객맞춤 서비스를 제공하여 고객의 만족을 증진시키며 지속적인 구매를 유도하는 것이다.

현재까지 웹 상의 추천 알고리즘 중 가장 성공률이 높은 것이 다음 <표 2>와 같은 입력자료선정, 이웃형성, 추천생

성의 단계로 나뉘는 CFR 알고리즘으로, 이는 한 활동고객에 대하여 다른 유사한 고객들인 이웃(neighborhood)들이 가장 선호한 상위 k 개 상품을 선별하여 추천한다[4].

〈표 2〉 CFR 알고리즘의 단계

단계	세부 알고리즘
① 입력자료 (DB)선정	고객-상품거래 행렬 $T_{(m \times n)}$ 를 랭크 k 를 갖는 축소행렬 $T_{(m \times k)}$ 로 선정하여 표준화(normalized)한다. 이때 m 명의 고객이 k 의 상품에 대한 메타(meta)거래를 갖으며 $k \ll m$ 이다.
② 이웃형성	$T_{(m \times k)}$ 로부터 고객간의 유사도 $sim(u, v)$ 를 계산한후, l 명의 이웃 고객군 N 을 형성한다.
③ 추천생성	현재 활동고객에 대하여 이웃 고객군 N 의 거래상품 중 k 개의 최빈 거래상품(또는 연관규칙)을 추천

[4]은 전통적인 연관규칙(association rule)과 협력적 여과규칙 중 전자상거래 데이터에 아이템 기반 협력적 여과규칙이 더 좋은 결과를 보임을 보였으며, 특히 추천상품의 개수가 늘수록 추천시스템의 성공률이 높아짐을 보였다.

2.2 MMDB

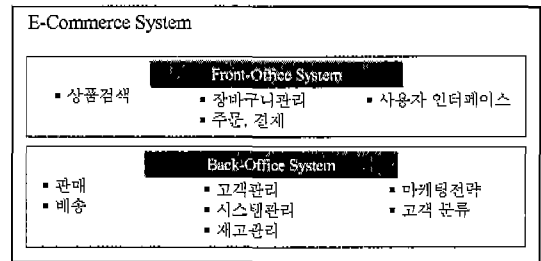
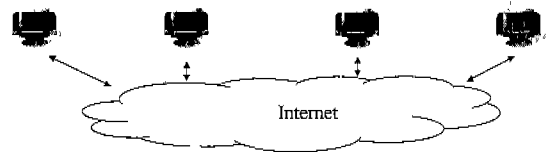
최근 기술개발의 성과로 인한 가격하락 때문에 그 활용 가치가 높아진 MMDB를 이용하여 데이터를 저장하고 처리한다. MMDB는 이동 통신 시스템, 공정 제어, 분산 가상현실망, 전자상거래등 많은 분야에서 데이터베이스에 대한 고성능 실시간 접근이 필요한 경우 도입하고 있는 기술이다 [6]. 기존 디스크 기반 DBMS는 디스크 접근의 오버헤드로 인하여 고성능 실시간 접근을 만족시키기 어렵기 때문이다. 전자상점의 규모가 대형화 되면서 많은 상품 중 고객이 원하는 상품을 제공하거나 보다 고객에게 맞는 개별화 서비스를 제공할 필요성이 생겨났으며, 이를 위한 다양한 고비용 연산이 필요하다. 디스크 기반 DBMS에 비하여 상대적으로 빠른 데이터 접근과 균일한 성능 분포 특성으로 인하여 전자상점의 최고 이웃고객의 협력적 여과처리와 추천을 제공하는데 기여할 수 있을 것으로 기대된다[9].

3. 시간에 민감한 상품추천 시스템

3.1 시스템 개요 및 구성

본 연구의 전자상점은 (그림 1)에서와 같이 고객에게 맞는 상품을 추천하여 거래가 이루어지는 개념적 공간인 전처리 시스템(Shop-front System)과 형성된 전자거래를 처리하는 공간인 후처리 시스템(Back-Office System)으로 구성된다.

먼저 전처리 시스템에서는 고객이 제품을 주문하기까지 고객이 사용할 수 있는 모든 기능인 상품검색, 장바구니 관



(그림 1) 개별화 상품 추천 전자상점

리, 주문, 결제 등을 지원하며, 고객이 직접 보고 사용하는 곳이므로 화면 인터페이스 설계가 매우 중요하다. 그리고 후처리 시스템에서는 판매, 구매, 재고, 배송, 제품, 고객, 시스템 관리 등 전자상점 운영 전반에 관련된 기능들이 있으며, 최근 들어서는 고객을 만족시키기 위한 고객관리부분이 매우 중요한 이슈로 떠오르고 있다.

고객관리를 위한 여러 서비스 중 고객의 정보를 이용한 고객과의 친밀감 있는 상호작용과 매출을 높이기 위하여 고객에게 적합한 상품을 추천하는 시스템의 연구 및 활용이 커지고 있다. 특히 본 연구에서는 상품 데이터베이스의 부분집합인 시간에 민감한 추천상품 데이터 기반으로 거래 데이터베이스로부터 이웃고객들의 협력적 추천에 의해 최

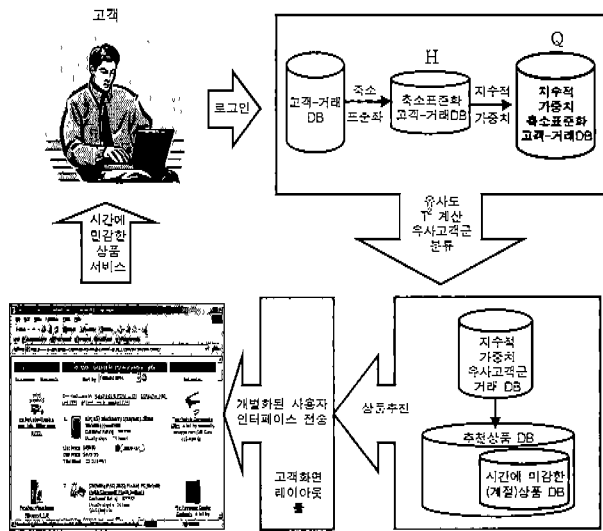
3.2 시간에 민감한 상품추천 시스템

현재까지 제안된 모든 협력적 여과 알고리즘은 모든 상품과 고객 거래데이터에 동일한 가중치를 주고 있다. 따라서 계절상품 또는 기획상품과 같이 시간에 따라 민감한 상품의 경우는 이와 같은 알고리즘이 효율적이지 못하고, 시간에 따른 고객 거래데이터의 가중치를 달리 주는 것이 알고리즘의 성공률과 고객의 만족을 높일 것이다.

즉, 최근의 고객 거래데이터에는 큰 가중치를 예전의 고객 거래데이터에는 지수적으로 감소하는 작은 가중치를 줌으로써, 유사고객군에게 최근에 인기가 있는 상품에 대한 추천확률을 높여주는 것이다. 따라서 지수적 가중치 협력적 여과 추천시스템은 특정 고객군에 대하여 최근접 이웃고객이 최근에 선택한 상품을 추천할 확률을 높여줌으로써 CRM을 지원하는 것으로, 시간에 민감한 계절상품이나 기획상품 등에 특히 효과적이다.

이와 같은 EWCFR 모듈은 다음의 (그림 2)과 같이 특정 활동고객이 전자상점에 로그인하면, 거래데이터베이스의 튜

플들은 지수적 가중치를 적용하여 표준화된다. 지수적 가중치가 적용된 거래 DB에 대하여 유사도가 계산되어 유사고객 클러스터 값을 출력하면, 이 값은 다시 협력적 추천시스템에 입력되어 추천상품 DB를 생성하여 레이아웃모듈을 통해 고객에게 추천된다. 이때 협력적 추천시스템은 최근 거래데이터일수록 지수적으로 큰 가중치를 부여하고 예전 거래데이터일수록 작은 가중치를 부여된 상품기반 협력적 추천상품을 계산한다. 이렇게 처리되어 가장 유사한 값을 갖는 추천상품을 정렬한 목록과 활동 고객의 기존 구매상품 목록과 비교하여 구매하지 않은 상품 목록에서 상위 k개의 상품을 추천하게 되는 것이다.



(그림 2) EWCFR 모듈 프로세스 구성도

본 절에서는 (그림 2)의 EWCFR 알고리즘 프로세스의 지수적 가중치 부여와 이를 이용한 유사도 계산에 대한 정리와 알고리즘을 정의하겠다.

고객 벡터 $\vec{c} = (c_1, c_2, \dots, c_m)$ 객체 콘텐츠인 상품 벡터 $\vec{p} = (p_1, p_2, \dots, p_n)$ 라고 하면, \vec{c} 은 \vec{p} 보다 정적(static)이며 대규모 전자상점의 경우 $m \geq n$ 가 될 것이다. 이때 \vec{p} 는 상품의 최하위 디렉토리 하에 있는 상품 벡터를 가정한다. 시스템에 로그인하여 다른 유사한 고객들이 선호하거나 구매를 한 상품을 추천 받게되는 활동 고객 A가 구매한 객체 콘텐츠 벡터를 \vec{b}_A 라 하면, $0 \leq \vec{b}_A \leq \vec{p}$, $A \in \vec{c}$ 이라 할 수 있다.

먼저 아이템(상품) 기반 모델링에 앞서 고객 DB의 협력 고객들의 해당 상품 구매데이터를 이용하여 표준화하여, 상점관리자에 의해서 미리 결정되어진 e개의 클러스터로 클러스터링하고 활동 고객 A가 속해있는 클러스터 $cluster_A$ 를 결정한다.

<정리>

고객거래 DB로부터 질의처리 얻은 i와 제품 j로 구성된 표준화 축소행렬 $H = \{h_{ij}\}$ 에 대하여 지수적 가중치를 적용한 협력적 축소행렬 $Q = \{q_{ij}\}$ 를 다음과 같이 정의할 수 있다.

$$Q_t = \Lambda H_t + (1 - \Lambda)Q_{t-1} \tag{1}$$

$$= \{q_{ij}\}, \quad i=1, 2, \dots, m, j=1, 2, \dots, n$$

$$t=0, 1, \dots$$

단, H_t : t시점의 유사거래고객 협력 축소행렬,

$\Lambda: (\lambda_1, \lambda_2, \dots, \lambda_m)$ 인 대각행렬, $0 < \lambda_j \leq 1$,

Q_0 : 유사고객 평균거래량 협력 축소행렬,

$Q_{(t-1)}$: (t-1)시점 지수적가중치 협력 축소행렬.

식 (1)의 Q_t 가 제공되면 고정된 임의의 시점 t에 대해서 식 (2)에 의해서 m명의 고객군에 대하여 최근접 이웃을 구할 수 있다.

$$SortMin(Sim_i^2) = SortMin(Q_i, \Sigma_i, Q_i) \tag{2}$$

단, Q_i 는 Q_t 행렬의 i고객에 대한 부분행렬이며, Σ_i 은 다음 식 (3)과 같이 구해지며, 만약 대각행렬 원소 λ 값이 모두 같다면 식 (4), 협력축소행렬의 크기가 충분히 크다면 식 (5)와 같이 계산된다.

$$\Sigma_{Q_i}(h, g) = \lambda_h \lambda_g \frac{1 - (1 - \lambda_h)(1 - \lambda_g)^t}{(\lambda_h + \lambda_g - \lambda_h \lambda_g)} \sigma_{h, g} \tag{3}$$

$$\Sigma_{Q_i}(h, g) = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}] \Sigma \tag{4}$$

$$\Sigma_{Q_i}(h, g) = \frac{\lambda}{2 - \lambda} \Sigma \tag{5}$$

단, $\sigma_{h, g}$: Σ 의 (h, g)번째 요소,

Σ : 협력축소행렬의공분산 행렬.

증명: 우선 식 (2) 좌항의 SortMin함수를 제외하여 고려해보자. 식 (1)의 Q_t 의 한 행인 고정된 고객 i에 대하여 식 (2)~식 (5)은 [1]에 의해 성립되고, 모든 고객들의 거래들은 서로 독립이므로 행렬 Q_t 에 대한 식 (1)의 증명이 완료된다. 다음 식 (2)는 고정된 임의의 시점 t에서 m명의 고객 유사도 $sim_i^2, t=0, 1, \dots$ 즉, 매트릭 Sim^2 이 얻어짐은 [1]에 의해 증명된다. 따라서 이 매트릭에 대하여 Min 함수를 사용하여 최근접 이웃을 구할 수 있다. □

위 정리에서 Sim^2 의 분포에 따른 알고리즘 평가기준의 제시는 본 논문의 범위를 벗어나며, λ_j 는 해당 상품의 거래데이터로부터 주기적으로 계산하여 얻을 수 있는 $0 < \lambda_j \leq 1$ 사이의 상수로써, 협력 축소행렬의 크기가 충분히 크지 않은 경우 유사고객의 협력 축소행렬을 활용하며 계산하기

위하여 1에 가깝게 주고, 계산하기 충분할 만큼 축소행렬의 랭크 값이 커지면 0에 가까운 가중치를 부여함으로써 최근 거래데이터에 높은 가중치를 부여하는 것이다.

다음은 앞의 <정리>에 의하여 지수적 가중치 엔트리 Sim_i^2 에 의한 추천 상품계산 알고리즘이다.

- [1] 최초의 구매고객($i=1$)인 경우, A 의 원소 값을 1로 주고 식 (1)을 계산하여 [3]단계로 간다.
- [2] 최초 고객이 아닌 경우, A 의 원소 값들을 비리 주기적인 추정에 의하여 얻어진 값(주로 0~0.3 사이)을 입력받아 식 (1)을 계산하고 다음 단계로 간다.
- [3] 고객수가 $m > kn$ 로 충분히 큰 경우 식 (5), A 의 원소값이 동일한 경우 식 (4), 그렇지 않은 경우 식 (3)을 계산한 후 다음 단계로 간다.
- [4] 지수적 가중치 유사도 Sim_i^2 를 계산한다.
- [5] $SortMin(Sim_i^2)$ 값에 의해 최고 근접한 이웃상품 상위 l 개를 추출하여 추천상품 벡터를 구성한다.
- [6] 추천상품 벡터에서 활동 고객 A 의 거래상품 벡터 \vec{p}_A 과 비교하여 중복되는 상품을 제외한 상위 k 개로 구성된 추천상품 벡터 \vec{r} 을 구한다. 즉, $\vec{r} \subset \vec{p}, r_i \neq p_A$.
- [7] 새로운 고객이 로그인하면 맨 첫 단계로 온다가 반복한다.

이와 같이 구성된 추천 콘텐츠 벡터에 의해서 고객이 전자상점에 로그인 하면 이미 MMDB에 저장되어 있던 유사 고객군 분류기를 통한 축소행렬 Q 에 의해, 해당 고객의 이웃이 계산되거나 연관규칙에 의해 최빈 구매상품이 검색된다. 이 객체 콘텐츠들이 해당 고객이 이미 구매한 상품은 제외한 상위 추천 콘텐츠를 취합한 컴포넌트 콘텐츠로 정리되어 고객에게 제공되는 것이다.

4. MMDB를 이용한 추천시스템

고객이 전자상점에 방문하였을 때 고객의 취향에 따라서 다른 레이아웃으로 상품 카탈로그를 제공하거나 관심 있는 물품 정보를 첫 페이지에 제공하거나 비교쇼핑을 할 수 있도록 정보를 저장해 둔다거나 등등 다양한 고객 개별화 서비스가 있다. 이러한 연산은 고객이 로그인 정보를 입력하고 첫 번째 페이지를 고객의 브라우저로 내려보내는 짧은 시간에 제공되어야 하므로 빠르게 처리하여야 한다. 따라서 빠른 서비스를 위하여 빈번히 요구되는 데이터를 핫데이터(hot data)로 정의하고, 이러한 데이터를 MMDB에 저장·관리하려는 하는 연구가 많이 진행되고 있다[9, 12].

이 방법은 핫데이터를 어떻게 정의하느냐에 따라 연산처리 속도나 비용이 결정되기 때문에 핫데이터의 정의방법이 중요하다. 본 연구에서는 핫데이터만 MMDB에 올리는 것이 아니라 모든 데이터를 MMDB에서 관리하므로 핫데이터

정의방법과 상관없이 연산처리 속도를 향상시킬 수 있다.

모든 고객을 (그림 2)과 같은 과정으로 유사고객분류기에 의하여 유사 고객군으로 분류한다. 즉, 고객이 전자상점에 방문하면 고객이 전자상점에서 행한 모든 트랜잭션을 기록한 데이터와 고객이 회원 가입시 제공한 정보를 가지고 데이터마이닝 틀에 의하여 고객의 선호를 파악한 후 유사고객군을 결정한다. 그리고 3.2절에서 소개한 바와 같이 유사고객군의 정보와 고객의 거래 데이터를 실시간으로 갱신하고 참조함으로써 고객이 원하는 상품을 예측하여 제시하게 된다.

이 과정에서 시스템은 대량의 고객 데이터베이스로부터 고객의 선호를 파악하기 위하여 검색 질의를 수행하며, 협력적 여과알고리즘이 상품데이터베이스를 참조하고 뽑아진 추천상품으로 사용자 인터페이스를 생성하여야 하므로 디스크 기반 데이터베이스로 다수의 고객에 대한 서비스가 만족할 만한 수준이 되기가 어렵다. 따라서 동시성제어, 데이터 접근속도, 질의처리 속도, 데이터 표현성능, 데이터 전송속도가 입증된 XML기반의 MMDB를 사용한다[9].

4.1 데이터 구조

전자상점에서 사용되는 모든 데이터는 데이터 자체에 메타데이터를 포함하고 있는 XML로 표현한다.

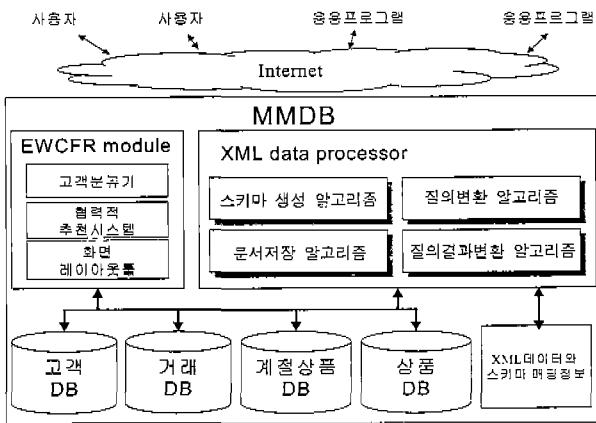
XML 데이터의 구조(Document Type Definition : DTD)가 DB의 스키마와 비슷하기 때문에 XML DTD를 파싱하여 XML데이터 처리를 위한 DB 스키마를 생성하는 연구가 많이 진행되었다[6]. 그러나 DTD는 그래프 형태와 유사하여 데이터베이스 구조와 완벽하게 일치하지 않으므로, XML 데이터를 객체-관계형 데이터베이스에 저장하려면 다음과 같은 오버헤드가 필요하다[9, 10].

첫째는 데이터를 저장하기 위하여 DTD를 DB 스키마로 변환해 주는 과정이 필요하며[9], 두 번째는 데이터를 검색하는데 사용되는 질의어가 서로 상이하기 때문에 XML질의어를 SQL형태로 변환하는 일이 필요하다. 마지막으로 DB에서 얻어낸 결과를 다시 XML로 변환해주어야 전자상점에 필요한 모든 서비스에 이용될 수 있다.

4.2 MMDB 구조 및 XML 데이터 처리 알고리즘

MMDB의 구조도는 (그림 3)과 같다. EWCFR 모듈에서 고객군을 분류하고, 협력적 추천시스템이 고객이 속한 고객군의 데이터로부터 상품을 추천한 후 레이아웃 틀로 고객에게 제공할 화면을 구성한다.

XML 데이터 처리모듈에서는 전자상점에서 다루는 모든 데이터의 구조에 따라 각 데이터베이스의 스키마를 생성한



(그림 3) MMDB 구조도

다. 스키마 생성알고리즘(그림 4)은 데이터 모델변환에 의한 매핑정보를 질의처리에 사용하기 위하여 저장한다. 문서 저장 알고리즘은 전자상점에 사용된 모든 XML데이터 파싱하여 DOM을 만들고, 스키마 생성 알고리즘에 의해 생성된 DTD와 스키마간의 매핑정보를 이용하여 클래스의 인스턴스로 저장한다(그림 5). 질의처리부분은 XML 질의어를 질의어 변환 알고리즘(그림 6)을 이용하여 객체-관계형 DB 질의어로 변환한다. 질의결과는 웹을 통해 사용자에게 전달되어야 하므로 질의결과변환 알고리즘(그림 7)에 의해서 다시 XML문서로 반환한다.

```

CREATE SCHEMA()
{
  if (token == ELEMENT)
  then if(Element_has_subelement())
    element_name = next_token();
    CREATE CLASS(element_name);
    while(token != '>')
    token = next_token();
    ADD_ATTRIBUTE(token, element_name); ;
  ;
  if (token == ATTLIST)
  then
    element_name = next_token();
    while (token != '>')
    token = next_token();
    att_name = Rename(element_name, token);
    ADD_ATTRIBUTE(att_name, element_name); ;
  ;
  CREATE MAPPING_INFORMATION();
  // 매핑정보를 MMDB에
  생성한다.
};
    
```

(그림 4) 스키마생성 알고리즘

스키마 생성 알고리즘을 거쳐 생성된 스키마는 의미적 제약사항을 표현하지 않고, XML데이터의 구조적 정보만 표현한 상태이다. [7]에서 제안한 방법에 의해 의미적 제약사항을 반영하여 스키마를 완성한다.

(그림 5)의 문서 저장알고리즘은 XML데이터를 DOM으로 변환한 후 스키마 매핑정보를 참조하여 XML 데이터를 데이터베이스에 저장한다.

```

DOC_INSERTION()
{
  doc = XML_Parser();
  create_DOM(doc);
  INSERT_INSTANCE();
}
INSERT_INSTANCE()
{
  Get_Mapping_Information();
  Insert_to_DB();
}
    
```

(그림 5) 문서저장 알고리즘

고객이 웹에서 작성한 상품 검색 질의는 (그림 6)의 질의 변환 알고리즘에 의해 SQL질의로 변환한 후 데이터베이스 질의처리에 전달되어야 한다. 즉, XML질의어를 파싱한 뒤 매핑정보를 이용하여 XML 엘리먼트를 DB 클래스로 변환한다. 만약 질의어가 구조정보에 맞지 않는 것이 입력되었을 경우 질의가 유효하지 않다는 메시지를 출력한 후 종료한다.

```

Query_Trans()
{
  query = XML_Parser(xql);
  if(validate(query)) {
    get_mapping_information();
    sql = translate_query();
    execute_query(sql);
  } else
  message("Query is not available");
}
    
```

(그림 6) 질의변환 알고리즘

질의가 유효할 경우 데이터베이스에서 질의처리에 질의를 전달하여 수행하도록 한다. 질의수행 후 DB에서 반환되는 질의결과는 다시 XML로 변환되어야 하므로 (그림 7)의 질의결과변환 알고리즘에 의하여 XML데이터로 변환한다.

```

Result_Constructor()
{
  resultset = Get_Query_result();
}
    
```

```

metadata= Get_Mapping_Information(resultset) ;
while(resultset.next()){
    construct_doc(resultset, metadata) ;
}
    
```

(그림 7) 질의결과변환 알고리즘

본 절에서는 효과적이고 빠른 상품추천시스템을 구현하기 위하여 데이터 자체에 메타데이터를 포함한 XML을 사용하였다. 또한 XML 데이터 처리의 오버헤드를 최소화하면서 고객의 개별화 서비스를 빠르게 처리할 수 있도록 MMDB에서 XML데이터를 처리할 수 있는 알고리즘을 설계하였다.

5. 결 론

본 논문에서는 전자상점의 추천시스템으로 가장 적합한 협력적 여과알고리즘을 전자상점의 주요 상품일 수 있는 인기상품, 계절상품, 특별기획상품과 같이 시간에 민감한 상품추천에 적합하도록 지수적 가중치론 적용한 EWCFR 알고리즘을 유도·제시하였다.

전자상점의 후처리시스템에 적용될수 있는 EWCFR 알고리즘은 개별고객마다 과거 거래 데이터보다는 최근의 거래데이터에 높은 가중치를 주어 계산한 후, 유사 고객군들이 협력적으로 상품을 추천하는 알고리즘을 적용한다. 이로써 전자상점의 매출에 막대한 영향을 미칠 수 있는 시간에 민감한 계절 또는 인기상품 거래정보를 보다 효율적으로 고려할 수 있으며, 최근 거래데이터의 신뢰도를 향상함으로써 고객들에게 추천한 상품의 구매성공률이 높아지는 것이다.

또한 전자상점의 빈번한 고객 접근으로 인한 데이터 처리의 오버헤드, 지수적 가중치 계산 비용과 협력적 추천알고리즘 연산비용의 부담을 줄여 고객별 빠른 개별화 서비스가 가능하도록 XML기반의 MMDB도 적용설계 알고리즘도 제공함으로써, 앞서 제시한 EWCFR 알고리즘의 실용성을 제고하고 후처리시스템의 빠른 계산처리로 인한 고객 만족과 전자상점 운영의 효율성을 추구하고자 하였다.

향후 연구로는 전자상거래 데이터로부터 연관규칙, 기존의 CFR 알고리즘, EWCFR 알고리즘을 각각 구현하고, 제1종 오류와 제2종 오류의 비율을 비교하는 성능평가 시뮬레이션과 현장의 사례연구이다.

참 고 문 헌

[1] 돈페퍼스, 마사로저스, “평생고객을 만드는 1대1 마케팅혁명,

CM비즈니스”, 1995.

[2] 이강수, “디지털 콘텐츠 엔지니어링 패러다임”, 2000 한국디지털컨텐츠학회 학술대회논문집, pp.65-75, 2000.
 [3] 한정혜, 임춘성, 이인경, “전자상점을 위한 멀티미디어 DB 인터페이스 구축”, 2000 한국디지털컨텐츠학회 학술대회논문집, pp.235-241, 2000.
 [4] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, “Item-based Collaborative Filtering Recommendation Algorithms,” www10 Conference, 2001.
 [5] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, “Analysis of Recommendation Algorithms for E-commerce,” Proceedings of ACM Conference, 2000.
 [6] B. Kao and H. Garcia-Molina, “An Overview of Real-Time Database Systems,” in Advances in Real-Time System(S. H. Son, ed.), ch. 19, Prentice Hall, 1995.
 [7] Dongwon Lee and Wesley W. Chu, “Constraints preserving Transformation from XML Document Type Definition to Relational Schema,” Proc. 19th Int’l Conf. on Conceptual Modeling (ER), Salt Lake City, Utah, 2000.
 [8] Gediminas A. and Alexander T., “User Profiling in Personalized Applications through Rule Discovery and Variations,” Proc. of the 5th ACM SIGKDD, 1999.
 [9] H. Garcia-Molina and K. Salem, “Main Memory Database Systems : An Overview,” IEEE Transactions on Knowledge and Data Engineering, 1992.
 [10] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl J., GroupLens : Applying Collaborative Filtering to Usenet News. 1997.
 [11] Lowry, C.A., Woodall, W.H., Champ, C. W., and Rigdon, S. E., “A Multivariate Exponentially Weighted Moving Average Chart,” Technometrics, 34, pp.46-53, 1992.
 [12] Pedro Trancoso, Josep Torrellas, “Cache Optimization for Memory Resident Decision Support Commercial Workloads,” Proceedings of the 1999 IEEE International Conference on Computer Design, 1999.



이 경 희

e-mail : khlee@trul.chungbuk.ac.kr

1999년 충북대학교 전자계산학과 졸업

(이학석사)

1999년~현재 충북대학교 전자계산학과

박사과정

2000년~현재 (주)엔솔래시닷컴 선임연구원

2001년~현재 극동정보대학 전산정보처리과 겸임교수

관심분야 : EC, 메인메모리 DB, 질의처리



한 정 혜

e-mail : hanjh@cje.ac.kr

1998년 충북대학교 전자계산학과 졸업
(박사)

1998년~1999년 연세대학교 산업시스템
공학과 포닥연구원

1999년~2001년 행정자치부 국가전문행정
연수원 통계연수부 전산교육
전임교수

2001년~현재 청주교육대학교 컴퓨터교육과 교수

관심분야 : 멀티미디어게임, EC, 데이터마이닝



임 춘 성

e-mail : leem@yonsei.ac.kr

1987년 서울대학교 산업공학과 (석사)

1992년 UC Berkeley 산업공학과 (박사)

1993년~1995년 미국 Rutgers University
산업공학과 조교수

1995년~현재 연세대학교 정보산업전공
부교수

관심분야 : 전자상거래, 기업정보화방법론