

# MMDB를 이용한 전자상거래 상품추천 시스템

김용기<sup>0</sup>, 이경희<sup>1</sup>, 한정혜<sup>2</sup>, 이충세<sup>3</sup>

충북대학교 전자계산학과<sup>0</sup>, 청주교육대학교 컴퓨터교육학과<sup>1</sup>

{wmm1203, khlee}@algo.chungbuk.ac.kr, hanjh@cje.ac.kr, csrhee@cbucc.chungbuk.ac.kr

## Recommendation System for E-Commerce using MMDB

YongGi Kim<sup>0</sup>, KyungHee Lee<sup>1</sup>, Jeonghye Han<sup>2</sup>, ChungSei Rhee<sup>3</sup>

Dept. of Computer Science, Chungbuk Nat'l Univ.<sup>0</sup>,

Dept. of Computer Education, Chongju Nat'l Univ. of Education<sup>1</sup>

### 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, 고객 정보 데이터베이스와 거래정보 데이터베이스로부터 연관규칙 등을 추출하여 사용한다. 그러나 시간의 흐름에 민감한 계절상품이나 특선상품과 같이 전자상점의 거래량에 크게 직결될 수 있는 것 등에도 같은 알고리즘을 적용한다면 추천성공률이 떨어질 것이다. 따라서 본 논문에서는 시간의 영향을 많이 받는 상품추천을 위하여, 최근 전자상점 추천시스템으로 효과적인 아이템 기반 협력알고리즘에 지수적 가중치를 적용하여 추천하는 알고리즘을 제안한다. 또한 이러한 추천시스템이 대용량의 고객데이터와 상품데이터에 대한 연산을 수행하고 다수의 고객에게 실시간으로 서비스를 제공하여야 하므로 MMDB를 활용한다.

### 1. 서 론

현재 세계는 인터넷에 의해 전자적 단일 상권으로 묶이는 본격적인 글로벌 전자상거래 시대로 돌입하고 있다. 실제로 국내외에서 이미 수많은 전자상점이 활발히 운영 중이며, 무한경쟁의 상거래 속에서 살아남기 위하여 원투원 마케팅 개념을 적용한 CRM(Customer Relation Management)을 도입하여 서비스를 제공하고 있다.

고객웨어전략의 가장 중요한 원칙이며 핵심은 파레토 원리-어떤 회사 비즈니스의 80%는 20%의 고객들로부터 창출된다-이다[1]. 실제 이 동통신 등 많은 회사는 CRM을 궁극적으로 이러한 로얄고객 확보 및 유지율 목표로 많은 비용을 들이고 있다. 이렇게 고객군을 분류하여 로얄고객을 관리하는 고객기반과 제품기반을 비교해보면, 인기상품의 매출이 전체 매출에 크게 영향을 주는 등 비슷하게 볼 수 있다. 또한 유행 및 시간에 민감한 제품이 매출에 영향을 준다 것을 고려하여 CRM을 하게 되면 효과적인 것이다. 그러자 아직까지 대부분의 전자상점들의 추천시스템은 고객과 상품을 대상으로 고정된 시점(time)하에서 연관규칙 또는 볼륨 계산하는 것이 거의 대부분이다.

따라서 본 논문은 전자상점의 매출에 영향이 큰 계절상품 또는 특별상품에 대한 추천시스템을 제안하고자 한다. 예를 들어, 계절상품에 어울리는 추천함에 있어 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(Collaborative Filtering) 알고리즘이 효과적으로 나타나는데, 즉 현재 로그인한 활동고객과 가장 유사한 이웃고객들의 거래 데이터베이스를 토대로 패턴을 계산하여, 결과에 따라 인기상품과 특별상품 또는 계절상품을 차별화 제시할 수 있으므로 효과적인 CRM이 이루어지게 되는 것이다. 이때 계절상품은 시간에 매우 민감하므로, 최근 이웃고객들의 거래데이터에 대한 가중치는 높게 그리고 이전 이웃고객들의 거래데이터에 대한 가중치를 낮게 적용하는 것이 효과적이다. 따라서 이를 위하여 [12]의 지수적 이동평균가중법(Exponentially Weighted Moving Average Method)을 적용한 협력적 여과 알고리즘을 제안하고자 한다.

그러나 이러한 상품들에 대한 고객의 반응(구매)은 시간에 매우 민

감하게 변화하므로, 고객거래 데이터베이스의 실시간 분석에 의한 추천을 해야 효과적인데, 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 데이터베이스를 이용하고자 한다.

본 논문에서는 고객 데이터베이스와 거래 데이터베이스를 분석하여 유사고객군으로 분류한 후, 해당 유사고객군의 고객데이터들을 CF 알고리즘을 통하여 특정고객이 구매할 가능성이 가장 크면서 시간에 민감한 계절상품 또는 인기상품을 추천하는 시스템을 제안하고자 한다. 이 추천 시스템은 고객이 전자상점에 방문한 시점부터 로그인할 때까지 개별화 서비스를 제공하며, 특히 로그인하는 시점에 고비용 절의가 많이 수행되기 때문에 빠른 응답을 제공하기 위하여 MMDB를 활용하고자 한다.

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

### 2. 관련연구

#### 2.1 전자상점 추천 시스템

현재 전자상점에 적용할 수 있는 상품 중개법으로는 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 나눌 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.

이러한 고객의 신상명세와 구매행동을 바탕으로 개별화된 셀링 기법을 통한 고객맞춤 서비스를 제공하여 고객의 만족을 증진시키며 지속적인 구매를 유도하는 것이다. [6]는 전통적인 연관규칙(association rule)과 협력적 여과규칙 중 전자상거래 데이터에 아이템 기반 협력적 여과규칙이 더 좋은 결과를 보임을 보였다. 현재까지 웹상의 추천 알고리즘 중 가장 성공률이 높은 것이 협력적 추천 알고리즘으로, 이는 한 활동고객에 대하여 다른 유사한 고객들인 이웃(neighborhood)들이 가장 선호한 상위 k개 상품을 선별하여 추천한다. 협력적 추천 여과 알고리즘은 입력자료선정, 이웃형성, 추천생성의 단계로 나뉜다[6].

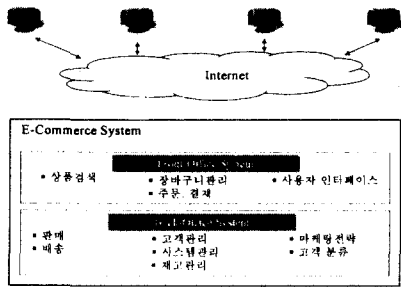
2.2 MMDB

최근 기술개발의 성과로 인한 가격하락 때문에 그 활용가치가 높아진 MMDB를 이용하여 데이터를 저장하고 처리한다. MMDB는 이동통신 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 어렵다. 따라서 메인 메모리가 대용량화되고 비용이 저렴해졌기 때문에 고객에게 개별화 서비스를 제공하기 위한 다양한 연산처리에 적합한 MMDB를 도입하였다. 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[8].

3. 시간에 민감한 상품추천 시스템

3.1 시스템 개요 및 구성

본 연구의 전자상점은 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. 즉, [그림 1]에서와 같이 먼저 전처리 시스템에서는 고객이 제품을 주문하기까지 고객이 사용할 수 있는 모든 기능인 상품검색, 장바구니 관리, 주문, 결제 등을 지원하며, 고객이 직접 보고 사용하는 곳 이므로 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 이유로 떠오르고 있다.



[그림 1] 개별화 상품 추천 전자상점

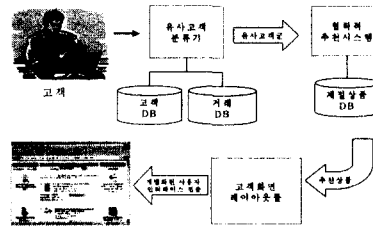
본 논문의 전자상점은 [그림 1]와 같이 구성되었는데, 상품 데이터 베이스로의 부분집합인 시간에 민감한 추천상품 데이터 기반으로 거래 데이터베이스로부터 이웃고객들의 협력적 추천에 의해 최고근접 상품을 추천하게 된다.

3.2 시간에 민감한 상품추천 시스템

[6]는 전자상거래 데이터에 아이템 기반 협력적 여과규칙이 더 좋은 결과임을 보였다. 그러나 현재까지 협력적 여과 알고리즘은 모든 상

품과 고객 거래데이터에 동일한 가중치를 주고 있으나, 계절상품 또는 기획상품과 같이 시간에 따라 민감한 상품의 경우는 시간에 따른 고객 거래데이터의 가중치를 달리 주는 것이 효율적이다. 즉, 최근에 인기가 있는 계절상품 또는 기획상품의 고객 거래데이터에는 큰 가중치를, 예전의 고객 거래데이터에는 지속적으로 감소하는 작은 가중치를 주는 것이다. 이러한 지수적 이동평균 가중치 협력적 여과 추천시스템은 특정 고객군에 대하여 최근접 이웃고객이 최근에 선택한 계절상품 또는 기획상품을 추천함으로써 CRM을 지원하는 것이다.

따라서 CFR 모듈은 다음 [그림 2]과 같이 특정 활동고객이 전자상점에 로그인하면, 관리자에 의해서 미리 선정되어 있는 해당 계절상품 또는 기획상품의 거래내역이 거래데이터베이스에서 추출되어 유사고객 분류기로 입력된다. 그러면 유사고객분류기는 해당 상품에 대한 구매총액과 인구통계적 유사도에 따라 고객들이 클러스터링하여 유사고객 클러스터 값을 출력하면, 이 값은 다시 협력적 추천시스템에 입력된다. 협력적 추천시스템은 최근 거래데이터일수록 지속적으로 큰 가중치를 부여하고 예전 거래데이터일수록 작은 가중치를 부여하여, 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] 프로세스 흐름도

고객 벡터  $\vec{c} = (c_1, c_2, \dots, c_m)$ , 객체 컨텐트인 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 디렉토리에 있는 상품 벡터를 가정한다. 시스템에 로그인하여 다른 유사한 고객들이 선호하거나 구매를 한 상품을 추천 받게되는 활동 고객 A가 구매 한 객체 컨텐트 벡터를  $\vec{p}_A$ 라 하면,  $0 \leq \vec{p}_A \leq \vec{p}, A \in \mathcal{A}$ 이라 할 수 있다.

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

그러면 활동 고객과 같은 클러스터에 속하는 협력 고객들의 상품구매 결과를 질의처리 추출하여, 고객 i와 제품 j로 구성된 축소행렬 H에 대하여 지수적 이동평균 가중치를 적용한 협력적 축소행렬 Q를 다음과 같이 정의할 수 있다.

$$Q_i = \lambda H_i + (1 - \lambda) Q_{i-1} \quad \text{〈식1〉}$$

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

단,  $H_i$ : t시점의 유사거래고객 협력 축소행렬,  
 $\lambda$ : 대각원소가  $(\lambda_1, \lambda_2, \dots, \lambda_m)$ 인 대각행렬,  
 $Q_0$ : 유사고객 평균거래량 협력 축소행렬,  
 $Q_{(t-1)}$ : (t-1)시점 지수적가중치 협력 축소행렬.

이때,  $\lambda_j$ 는 해당 상품의 거래데이터로부터 주기적으로 계산하여 얻을 수 있는  $0 < \lambda_j \leq 1$ 사이의 상수로서, 현재 고객에게 제품을 추천하기 위한 협력 축소행렬의 크기가 충분히 크지 않은 경우, 인구통계적 유사 고객의 협력 축소행렬을 활용하며 계산하기 위하여 1에 가깝게 주고

계산하기 충분한 만큼 축소행렬의 랭크 값이 커지면 0에 가까운 가중치를 부여함으로써 최근 거래데이터에 높은 가중치를 부여하는 것이다.

<식1>에 대하여 [12]로부터 다음 <식 2>에 의해서 유사도  $T_i^o$ 로부터 최근접 이웃상품을 구할 수 있다.

$$T_i^o = Q_i \Sigma_o^{-1} Q_i \quad \text{<식2>}$$

이때  $\Sigma_o$ 는 다음 <식3>과 같이 구해지며, 만약 대각행렬 원소  $\lambda$  값이 모두 같다면 <식4>, 협력축소행렬의 크기가 충분히 크다면 <식 5>와 같이 주워진다.

$$\Sigma_o(h, g) = \lambda_h \lambda_g \frac{1 - (1 - \lambda_h)(1 - \lambda_g)^t}{(\lambda_h + \lambda_g - \lambda_h \lambda_g)^t} \sigma_{h, g} \quad \text{<식3>}$$

$$\Sigma_o(h, g) = \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}] \Sigma \quad \text{<식4>}$$

$$\Sigma_o(h, g) = \frac{\lambda}{2 - \lambda} \Sigma \quad \text{<식5>}$$

단,  $\sigma_{h, g}$ :  $\Sigma$ 의  $(h, g)$ 번째 요소,  
 $\Sigma$ : 협력축소행렬의공분산 행렬.

다음은 지수적 가중치 엔트리  $T_i^o$ 에 의한 추천 상품계산 알고리즘이다.

- [1]최초의 구매고객( $i=1$ )인 경우,  $A$ 의 원소 값을 1로 주고 <식1>을 계산하여 [3]단계로 간다.
- [2]최초 고객이 아닌 경우,  $A$ 의 원소 값을 미리 주기적인 추정에 의하여 얻어진 값을 입력받아 <식1>을 계산하고 다음 단계로 간다.
- [3]고객수가  $m > kn$ 로 충분히 큰 경우 <식5>,  $A$ 의 원소값이 동일한 경우 <식4>, 그렇지 않은 경우 <식3>을 계산한 후 다음 단계로 간다.
- [4] 지수적 가중치 유사도  $T_i^o$ 를 계산한다.
- [5]  $T_i^o$ 값에 의해 최고 근접한 이웃상품 상위  $l$ 개를 추출하여 추천 상품 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,  $\vec{r} \subset \vec{p}_A, r_i \neq p_{Ai}$ .
- [7]새로운 고객이 로그인하면 맨 첫 단계로 올라가 반복한다.

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

#### 4. MMDB를 이용한 전자상거래 추천시스템

메인메모리의 가격하락으로 빠른 응답시간을 요구하는 애플리케이션에서 필요한 데이터를 MMDB에 저장함으로써 원하는 성능을 얻어내고 있다. MMDB는 디스크 기반 데이터베이스에 비하여 훨씬 빠른 응답시간과 트랜잭션 처리속도를 나타내기 때문에 실시간 처리가 필요한 제어시스템이나 통신시스템에 많이 사용되어왔다. 최근에는 많은 전자상점에서 고객의 만족도를 높이기 위하여 고객이 전자상점에 방문하였을 때 질 높은 서비스를 빠르게 제공하기 위하여 빈번히 요구되는 hot 데이터[9]를 MMDB에 저장관리하려는 연구가 많이 진행되고 있다[12].

본 연구에서는 [그림 2]과 같은 과정으로 CRM을 구현한다. 모든 고객을 유사고객분류기에 의하여 유사고객군으로 분류한다. 고객이 전자상점에 방문하면 고객이 전자상점에서 행한 모든 트랜잭션을 기록한 데이터와 고객이 회원가입시 제공한 정보를 가지고 데이터마이닝 툴에 의하여 고객의 선호를 파악한 후 유사고객군을 결정한다. 그리고 3.2절

에서 소개한 바와 같이 유사고객군의 정보와 고객의 거래 데이터를 실시간으로 갱신하고 참조함으로써 고객이 원하는 상품을 예측하여 제시하게 된다. 이 과정에서 시스템은 대량의 고객데이터베이스로부터 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 동시성 제어, 데이터 접근속도, 질의처리 속도, 데이터 표현성능, 데이터 전송속도가 입증[9]된 메인메모리 데이터베이스를 도입한다.

#### 5. 결 론

본 연구에서는 전자상점의 추천시스템으로 가장 적합한 협력적 여과 알고리즘을 전자상점의 주요 상품일 수 있는 인기상품, 계절상품, 특별 기획상품과 같이 시간에 민감한 상품추천에 적합하도록 지수적 이동평균 가중법을 적용하여 유도하였다. 즉, 개별고객에 최근의 거래데이터에 높은 가중치를 주어 계산하여 추천해 주는 협력적 추천 알고리즘을 제안하였으며, 전자상점의 빈번한 고객 접근으로 인한 데이터 처리의 오버헤드, 협력적 추천알고리즘 연산비용, 고객별 빠른 개별화 서비스 등에 적합한 MMDB도 도입하였다.

따라서 전자상점의 매출에 영향을 크게 미치는 시간에 민감한 상품에 대하여, 보다 많은 고객들에게 추천한 상품의 구매성공률이 높도록 하여 고객만족과 전자상점 운영의 효율성을 추구하였다.

향후 연구로는 실제 평가문제에 대하여 연관규칙과 협력적 여과알고리즘의 CFR을 각각 구현하고, 제1종오류와 제2종오류의 비율을 비교하여 성능평가가 될 것이다.

#### 참 고 문 헌

- [1] 돈페퍼스, 마사로저스, (1995), "평생고객을 만드는 1대1 마케팅 혁명, CM비즈니스".
- [2]박주환, 이순미, 이재호(2001), "교육환경에서 고객관계관리 기술의 활용", 2001 동계 학술발표 논문집 6권 1호, 109-120.
- [3]이강수(2000), "디지털 콘텐츠 엔지니어링 패러다임", 2000 한국디지털콘텐츠학회 학술대회 논문집, 65-75.
- [4]한정혜, 임춘성, 이인경(2000), "전자상점을 위한 멀티미디어 DB 인 터페이스 구축", 2000 한국디지털콘텐츠학회 학술대회 논문집, 235-241.
- [5]Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl(2001), "Item-based Collaborative Filtering Recommendation Algorithms", www10 Conference.
- [6]Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl(2000), "Analysis of Recommendation Algorithms for E-commerce", Proceedings of ACM Conference.
- [7]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.
- [8] H. Garcia-Molina and K. Salem,(1992) "Main Memory Database Systems : An Overview," IEEE Transactions on Knowledge and Data Engineering, Dec.
- [9] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl J., GroupLens: Applying Collaborative Filtering to Usenet News. 1997.
- [10] Lowry, C.A., Woodall, W.H., Champ, C. W., and Rigdon, S. E. (1992); "A Multivariate Exponentially Weighted Moving Average Chart", Technometrics, 34, 46-53.
- [11] Pedro Trancoso, Josep Torrellas, (1999) "Cache Optimization for Memory-Resident Decision Support Commercial Workloads", Proceedings of the IEEE International Conference on Computer Design.