

연관규칙과 협업적 필터링을 이용한 상품 추천 시스템 개발

김지혜[†], 박두순^{††}

요 약

전자상거래가 급속도로 발전함에 따라 고객들의 행동 패턴을 어떻게 발견하느냐와 웹 마이닝 기술을 사용하는 것에 의해 어떻게 상거래를 지능화 할 것인가에 대한 연구가 진행되고 있다. 현재까지 개인화와 상품 추천 시스템을 만들기 위해 가장 성공적이고 가장 넓게 사용되는 기술은 협업 필터링 방법이다. 그러나 협업 필터링 방법은 특정 수 이상의 아이템에 대한 평가가 필요하다는 문제를 가지고 있다. 또한, 기존의 연관 규칙 기법은 개인별 사용자의 성향을 반영하지 못하는 단점을 가지고 있다. 본 논문에서는 개선된 Apriori 알고리즘을 이용하고, 아이템들 간에 상호 관계를 가진 협업 필터링 방법을 사용하여 사용자 성향이 반영된 상품 추천 시스템을 개발하였다.

키워드 : 연관규칙, 협업적 필터링, 개인화, 전자상거래, 상품추천시스템

Development of the Goods Recommendation System using Association Rules and Collaborating Filtering

Ji-Hye Kim[†], Doo-Soon Park^{††}

ABSTRACT

As e-commerce developing rapidly, it is becoming a research focus about how to find customer's behavior patterns and realize commerce intelligence by use of Web mining technology. One of the most successful and widely used technologies for building personalization and goods recommendation system is collaborating filtering. However, collaborative filtering have serious data sparsity problem. Traditional association rule does not consider user's interests or preferences to provide a user with specific personalized service. In this paper, we propose an goods recommendation system, which is integrated an collaborative filtering algorithm with item-to-item corelation and an improved Apriori algorithm. This system has user's interests or preferences ro provide a user with specific personalized service.

1. 서 론

전자상거래란 전자적 방식을 이용하여 가상공간에서 이루어지는 제반 거래행위이다. 전자상거

래는 정보통신 기술과 정보시스템 개발 기술의 발전으로 나타나는 새로운 사회제도이며 새로운 문화이다. 거래인증, 대금결제, 소비자보호, 지적 소유권보호 등에 관해 새로운 정책들이 수립되었고 시장은 전 세계로 확대되어 나갈 것이다.

전자상거래에서 중요한 성공요건 중 하나는 고객과의 효과적인 커뮤니케이션의 실현이다. 하지만 현재 전자 상거래에서 고객과의 효과적인 커

[†]학생회원: 순천향대학교 전산학과 석사과정
^{††}정 회원: 순천향대학교 정보기술공학부 교수(교신저자)
논문접수: 2005년 11월 24일, 심사완료: 2006년 1월 5일
* 본 논문은 2005학년도 순천향대학교 학술연구조성비 및
반연구과제에 의하여 연구되었음

뮤니케이션을 이루지 못하므로 추천 시스템과 같은 다른 차원의 접근방식이 필요하다[1].

상품 추천 시스템 개발 방법으로 로그파일분석 방법은 추천의 정확도에서 한계를 나타내고 있고, 최근에 데이터마이닝이나 CRM 시스템 등을 이용한 상품 추천 시스템을 사용하고 있지만, 개발과 관리의 어려움으로 더 쉽고 이용이 간편하며 비용이 적게 드는 소규모 개인화(Personalization) 서비스에 대한 연구가 활발하게 진행되고 있다[2]. 개인화를 통해서 웹 사이트 운영자는 사용자에 관한 자료를 얻고 사용자의 지속적인 이용이나 구매를 얻어낼 수 있게 되며 자신에게 가장 알맞은 정보를 편리한 방법으로 얻을 수 있다. 개인화 기술로는 규칙 기반(Rules-Based)과 협업 필터링(Collaborative Filtering), 학습에이전트(Learning Agent) 기법[4] 등이 있으며, 협업 필터링 기법이 현재 가장 많이 사용되고 있다. 하지만 협업 필터링 기법은 초기 평가자 문제와 이용자가 평가한 데이터가 희박함의 문제가 있다.

본 논문에서는 고객 데이터베이스에서 판매와 관련된 정보에 데이터마이닝의 연관규칙 중 가장 대표적인 Apriori 알고리즘에 지역적인 정보를 이용한 개선된 Apriori 알고리즘[3]을 적용하였다. 그러나 이 방법은 개인성향을 반영하지 못하는 문제점이 있다. 이를 보완하고자 협업 필터링 기법을 적용시켰으며 이로 인해 어떤 형태든 정보 검색이 용이하고 이용자 기호에 맞는 추천을 할 수 있으며 이용자의 검색 없이 자동 추천이 가능하다. 그러나 이 방법도 이용자가 평가한 데이터가 희박함의 문제가 발생하므로 항목들에 가장 높은 상관관계의 상품을 추천하는 항목 대 항목 상관관계를 적용하여 보완하였다.

본 논문의 구성은 2장에서는 연구배경 및 관련 기술을 서술하였고, 3장에서는 상품 추천 시스템을 설계하였으며, 4장에서는 구현 된 내용을 서술하였다. 마지막으로 5장에서는 결론을 맺는다.

2. 연구배경 및 관련 기술

상품 추천 시스템은 고객들에게 추천 상품 리스트를 만들어 고객들이 구매 가능성이 있는 상

품을 쉽게 찾도록 도와주는 개인화된 정보 필터링 기술이다. 본 절에서는 웹마이닝, 개인화, 연관규칙과 Apriori 알고리즘을 기술한다.

2.1. 웹 마이닝

현재 상품 추천 시스템을 운영하는 가장 대표적인 Amazon.com, CDNOW, eBay, Moviefinder.com 등의 인터넷 업체에서는 사용자의 구매기록, 즉 취향에 맞는 항목 및 정보를 사용하여 사용자에게 다양한 방법으로 추천해주고 있다.

웹 마이닝은 웹 문서와 서비스들로부터 자동적으로 정보를 발견하고 뽑아내기 위해 데이터 마이닝 기법들을 사용한다. 웹 마이닝은 웹 콘텐츠 마이닝(Web Content Mining), 웹 구조 마이닝(Web Structure Mining), 웹 사용자 마이닝(Web Usage Mining)으로 분류한다[4]. 웹 콘텐츠 마이닝은 웹 내용/데이터/문서에서 유용한 정보를 탐색하는 것을 의미하며 정보탐색의 관점에서 엠파스의 자연언어 검색 같은 자연어 처리시스템이나, 개인화된 웹 에이전트에 이용될 수 있고, 데이터베이스 측면으로는 데이터베이스의 축적된 데이터 중에서 원하는 자료를 쉽게 찾을 수 있게 해주는 구조화된 질의 언어를 개발하는데 이용된다. 웹 구조 마이닝[5]은 웹 사이트와 웹 사이트의 하이퍼 링크 정보를 데이터 마이닝 과정을 통해 구조화하고, 표준화 시키는 프로세스를 말하며 링크된 하이퍼 링크의 위상(topology)에 근거하여, 범주화, 군집화를 수행함으로써 유사한 웹 페이지들을 묶거나, 서로 다른 사이트 간의 관계와 유사성에 관한 정보를 얻는데 유용하다. 웹 사용자 마이닝[6]은 웹 로그나 쿠키 정보, 콘텐츠 데이터나 실제 사용자가 마우스 클릭, 스크롤 하는 것의 모든 데이터를 사용한다. 이런 데이터들을 사용하여 사용자의 접속패턴을 알아내고, 여기서 발견된 접속패턴이나 관심부분에 따라서 개인화 맞춤 서비스가 가능하다. 웹 사용자 마이닝은 가장 많이 사용되는 기술이며 본 논문에서 사용하였다.

웹 사용자 마이닝 처리과정은 전처리, 패턴발견, 패턴분석 등 크게 세부분으로 나눌 수 있다. 전처리 단계는 다양한 데이터 자원내의 사용 패턴 정보를 추상화 시키는 변환 과정으로 허용 가능한 데이터의 불완전성 때문에 매우 정교하고 복잡 처리절차를 요구한다. 패턴 발견 단계에서는 기존의 데이터 마이닝에서 쓰이는 방법론이나 알고리즘을 기반으로 한 패턴 탐색이 이루어지게 된다. 패턴 분석 단계에서는 흥미롭지 않은 규칙들을 잘라 내거나 원하는 패턴 형식을 발견하기 위해 질의 메커니즘과 OLAP 과 같은 기법 등을 사용할 수 있다. 패턴 발견을 위해 사용되는 방법론으로는 기계학습, 통계 기법, 연관규칙, 군집화, 분류화 등이 있으며, 발견된 패턴은 시스템 성능 향상, 사이트 유지 및 보수, 비즈니스 인텔리전스(Business Intelligent), 사용자 특성화 등에 활용 할 수 있다[7].

2.2. 개인화

웹 사이트를 개인화하기 위한 일반적인 기법들 중 규칙기반 기법은 사용자들에게 몇 가지 질문들을 한 이후에 대답한 답에 대해 내용을 전달하는 것이다. 여기서는 제공하는 질문을 통해 사용자들을 구분하고 개개인을 구별하기 위한 목적으로 사용된다는 점에서 개인화와는 큰 차이가 있다. 일반적으로 사용자의 인구통계학적 정보나 사이코 그래픽 정보를 사용자 확인의 중요한 요소로 사용되며 높은 수준의 개인화를 달성하기 위해 추가적으로 사용자의 선호도 정보를 사용해야 한다. 적절하게 선호도 정보를 제시하고 얻어진 자료를 바탕으로 사용자를 구분하기 위해서는 웹에서 가지고 있는 내용과 웹의 목적에 대한 체계적인 정의 작업이 필요하다. 이런 것이 규칙기반에서 가장 시간이 오래 걸리고 적용이 어려운 점이다.

협업필터링 기법은 추천 시스템에서 가장 흔히 쓰이는 기법이며, 사용자가 자발적으로 제공한 정보를 사용하여 사용자를 비슷한 선호도를 가진 집단으로 나누어 그 집단 내에서 서로에게 추천하는 방식을 사용한다. 집단 형성의 과정과 교차

추천의 과정은 자동으로 이루어지게 되고 사용자가 처음 사용하는 경우라 하더라도 충분한 자료가 축적되어 경우 사용자에게 즉시 서비스가 가능하기 때문에 웹 사용자들의 즉각적인 성향을 만족시켜줄 수 있는 방법이다. 또한 본질적으로 사용자의 개인정보를 공개하지 않아도 서비스 제공이 가능하기 때문에 최근의 개인정보보호를 우선시 하는 익명 개인화 추세에도 잘 맞는 방법이다.

그러나 새로운 항목이 추가되면 누군가가 선호도를 입력하거나 이용하기 전에 그 항목을 추천할 수 없는 초기평가자문제가 있다. 또 대부분의 영역에서 다수의 항목에 대해 모든 이용자가 평가할 수 없기에 이용자-항목 행렬의 평가 값들은 매우 희박한 상태가 되는 희박함의 문제가 있다.

학습에이전트 기법은 사용자의 웹상에서 활동을 관찰하고 사용자가 어떤 내용에 관심을 가지고 있는지 판단하여 사용자에게 알맞은 내용을 전달하도록 하는 것을 말한다. 사용자의 웹에서의 행동 중 중요하게 사용되는 것은 특정한 페이지를 보는 시간, 인쇄한 페이지, 전자상거래 사이트의 경우에는 구매한 상품과 쇼핑 카트에 넣은 상품 등을 들 수가 있으며 이러한 사용자의 행동을 관찰하기 전에 선행되어야 할 것은 웹 페이지의 내용들에 대한 정보가 일정한 기준에 따라 분류되어 있어야 한다는 것이다.

전자상점들은 개인화를 위해 개인화 되지 않은 추천(Non-Personalized Recommendation), 속성 기반 추천(Attribute-based Recommendation), 사람 대 사람 상관관계(People-to-People Correlation), 항목 대 항목 상관관계(Item-to-Item Correlation)의 4가지 추천 기술을 사용하는 데 개인화되지 않은 추천기술은 고객들의 의견을 종합해 평균적으로 제품을 추천해주는 방법으로 모든 고객이 동일한 추천을 받게 된다. 속성 기반 추천은 상품의 속성을 기반으로 제품을 추천해주는 기술이다. 사람 대 사람 상관관계는 상품에 대한 평가와 구매를 근거로 유사한 고객을 찾아 유사한 고객이 구매한 상품을 추천하는 기술이다. 항목 대 항목 상관관계는 고객이 관심을 표명한 작은 규모의 제품을 바탕으로 가장 높은 상관관계가 있는 상품을 추천하는 기술

이다.

2.3. 연관규칙과 Apriori 알고리즘

데이터들의 빈도수와 동시 발생 확률을 이용하여 한 항목들의 그룹과 다른 항목들의 그룹 사이에 강한 연관성이 있음을 밝혀 주는 기술이다. 발견된 연관성 규칙을 이용함으로써 진열대에 상품을 진열하는 일, 매출을 증가시키기 위해 카탈로그를 구성하는 일, 패키지 상품을 구성하는 일과 같은 기업업무에 도움을 줄 수 있다. 이러한 일은 마케팅에서 고객의 장바구니 안에 들어 있는 품목간의 관계를 알아본다는 뜻에서 시장바구니 분석이라 한다. 수많은 항목들 사이에서 의미 있는 연관성을 찾아내기 위해서는 연관규칙을 비교할 수 있는 기준이 필요하다. 이를 위해 데이터 마이닝에서는 지지도, 신뢰도, 리프트의 3가지 개념을 사용한다. 연관성을 분석하기 전에 미리 지지도, 신뢰도, 리프트의 최소기준치를 정하여 두고, 이 기준치를 초과하는 규칙만을 생성하게 함으로써 의미 있는 규칙만을 추출하는 것이다. 지지도는 전체 거래 중 관련성이 있다고 판단되는 항목들을 포함하는 확률이다. 신뢰도는 항목 A를 구입할 경우와 항목 A, B를 동시에 구매할 경우를 비교함으로써 A, B간의 상대적인 연관성을 측정한다. 리프트는 항목 A를 구입하는 것을 전제로 하고 항목 B를 구입하는 것이 얼마나 가능성이 있는지를 나타낸다.

또한 빈발항목(Large Item set) 집합을 생성하는 방법으로 가장 대표적인 Apriori 알고리즘이 있으며, 연관규칙을 구현하는 대표적인 알고리즘이다. <표 1>은 Apriori 알고리즘이다.

<표 1> Apriori 알고리즘

```

Ck: 크기 k의 후보항목집합
Lk: 크기 k의 빈발 항목집합
L1 = {빈발 항목};
for (k = 1; Lk != k++ ) do begin
Ck+1 = Lk로부터 생성되는 후보 항목들
for each transaction t in database do
    increment the count of all candidates
    in Ck+1 that are contained in t
    
```

```

Lk+1 = 최소신뢰도를 가진 Ck+1 안
    에 있는 항목들
end
return Uk Lk
    
```

C_k 는 Size가 K 인 후보 항목 집합(Candidate item set)으로서, L_{k-1} 로부터 얻어진다. 예를 들면 $L_2 = \{1,2\}, \{3,4\}$ 라고 하면 L_2 로부터 얻어지는 모든 C_3 은 $\{1,2,3\}, \{1,2,4\}, \{2,3,4\}$ 가 된다.

L_k 는 Size가 K 인 빈발 항목 집합으로써, C_k 로부터 얻어진다. 특정 값의 지지도 이상을 만족하는 집합들을 의미한다. 알고리즘 L_1 (크기가 1이고 최소 지지도를 만족하는 항목들의 집합)부터 시작하여, 크기를 하나씩 증가하는 방식으로 L_k 들을 찾아내는 과정이다.

더 이상의 L_k 를 만들어낼 수 없을 때 프로그램은 종료된다. 즉, 더 이상의 후보 항목 집합과 빈발 항목 집합을 만들어낼 수 없으면 프로그램은 종료된다.

트랜잭션 데이터베이스는 중복을 허용하지 않고 순수한 집합의 개념을 확장해 트랜잭션과 다른 모든 항목집합들 내에 존재하는 항목들은 이미 정렬돼 있는 것으로 가정해야한다.

트랜잭션 데이터베이스에서 연관규칙을 찾는다는 것은 사용자가 명시한 최소지지도와 최소신뢰도보다 큰 지지도와 신뢰도를 갖는 항목집합을 찾는 것이라고 말할 수 있고 최소지지도보다 큰 지지도와 갖는 집합들을 빈발항목이라고 말한다. 이런 경우 항목들의 집합 X 를 빈발 항목이라고 한다[8].

3. 연관규칙과 협업 필터링을 이용한 상품 추천 시스템 설계

본 논문에서 설계한 상품 추천 시스템은 웹 마이닝 기법 중 연관규칙을 사용하였고 개인화를 위해 협업 필터링 기법을 이용하였다. 연관규칙은 대용량의 데이터베이스에서 항목간의 지지도와 신뢰도에 근거하여 규칙을 발견하는 특징을 가지고 있다. 하지만 개인성향을 무시하고 항목간의 연관성을 근거로 항목을 추천하게 된다. 이

발생하는 특성이 있을 때 이를 고려한 빈발구간 항목집합 생성 방법이다. 이 알고리즘은 많은 수의 후보 항목집합을 생성하지 않고, Apriori-gen이라는 새로운 후보항목집합 생성전략을 개발하여, 후보 항목 집합의 수를 줄이는데 성공적이다.

<표 2> 개선된 Apriori 알고리즘

```

- 개선된 Apriori 알고리즘
// User specifies Minimum Support (Smin)
// Create f(lq) through search DB
FL = Ψ
for (k=1 ; Lq≠Ψ k++) do begin
  Max_lq = MAX(f(lqi), (not tagged,
                                1≤i≤n)
  fk merge lqi
    CALL Gen_FL
  FL = U fk// Answer
  Max_lq = 0
  lqs들의 합이 의미있는 구간안에 있는지 없는지를 체크
  if meaningful
    then divide intervals to two section
      FL = ▼ FLi, FLj
    else
      lqi = tagged (1in)
  endif
end
end

- Apriori-Gen 함수

Function Gen_FL
for (j=1; Max_lq ≥ Smin, j++) do begin
case 1 : lqti-j, and lqti+j are not tagged
if (f(lqi-j) + f(lqi+j)) ≤ (Smin - Max_lq)
then Max_lq = Max_lq + f(lqi-j) + f(lqi+j);
fk merge lqi-j, lqi+j lqti-j, lqti+j = tag
else if (f(lqi-j) ≤ f(lqi+j) and
(Smin - Max_lq) ≤ f(lqi-j))
then Max_lq = Max_lq + f(lqi-j);
fk merge lqi-j lqti-j = tag
else Max_lq = Max_lq + f(lqi+j);
fk merge lqi+j lqti+j = tag
endif
endif
case 2 : lqti-j is not tagged, lqti+j tagged
Max_lq = Max_lq + f(lqi-j) ;
fk merge lqi-j / lqti-j = tag
case 3 : lqti-j is tagged, lqti+j not tagged
Max_lq = Max_lq + f(lqi+j) ;
fk merge lqi+j / lqti+j = tag
case 4 : lqti-j and lqti+j is tagged
return// fk is not large
end
Return
    
```

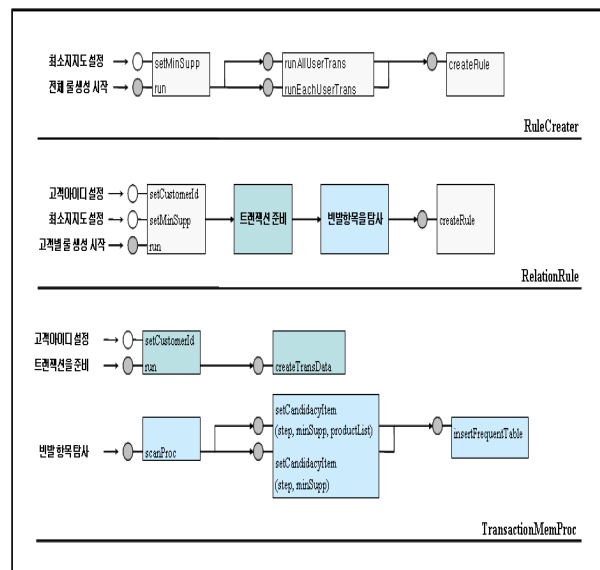
본 논문에서 주어진 데이터베이스에서 연관규칙을 찾는 것은 <표 3>과 같이 두 단계이다.

<표 3> 연관규칙을 찾는 두 단계

단계 1.
명세된 최소지지도(0.5)를 만족하는 모든 항목 집합들을 찾는다. 하나의 항목 집합에 대한 지지도는 해당 항목 집합을 포함한 트랜잭션의 수로 생각할 수 있으며, 최소 지지도를 만족하는 항목 집합을 빈발 항목 집합이라 하며, 빈발 항목집합을 찾기 위해서 개선된 Apriori 알고리즘을 사용한다.
단계 2.
데이터베이스로부터 연관규칙을 생성하기 위해 빈발 항목 집합을 사용하게 되는데, 모든 빈발 항목 집합에 대해서 공집합이 아닌 모든 부분 집합을 찾아야 한다. 이러한 각각의 부분 집합에 최소신뢰도 이상이면 연관규칙을 찾을 수 있다.

2.6. 상품 추천 시스템 엔진 모듈 설계

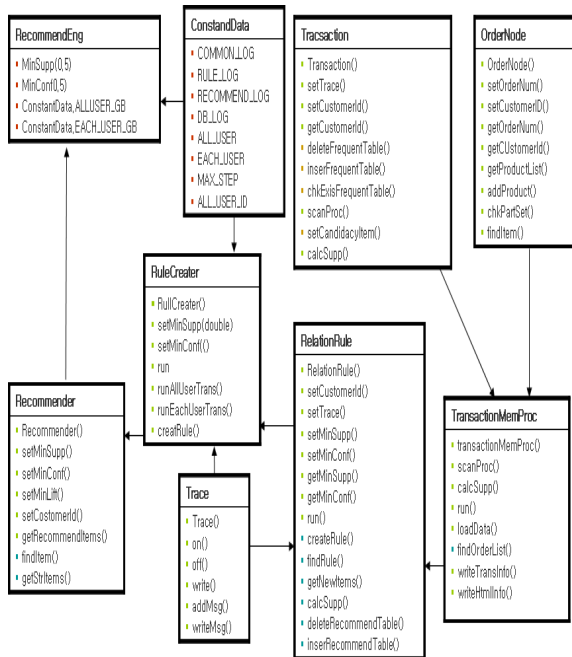
본 논문의 상품 추천 시스템 엔진은 자바 빈즈를 활용하였으며, (그림 3)은 상품 추천 시스템 엔진이다. 상품 추천에 직접적인 관련이 있는 클래스는 그림에 있는 RuleCreator, RelationRule, TransactionMemProc의 세 클래스이다.



(그림 3) 상품 추천 프로그램 설계

TransactionMemProc 클래스는 후보 항목 집합을 생성하는 등의 트랜잭션 데이터와 관련된 일을 담당한다. TransactionMemProc 클래스는 트랜잭션 데이터를 메모리에 세팅하고 관리하는 등의 일을 한다. 상품 추천 규칙 생성을 위해 트랜잭션 데이터에서 지지도와 신뢰도를 계산해 데이터베이스에 저장한다. 웹 관리자가 최소 지지도와 최소 신뢰도를 명시해 서버로 요청하면 명시된 지지도와 신뢰도를 만족하는 항목집합을 결과로 관리자에게 전송하는 루틴을 가지고 있다. 트랜잭션 데이터는 주문정보 테이블에서 트랜잭션에 필요한 정보만 추출하여 생성된다. RelationRule 모듈에서는 트랜잭션에서 빈발 항목을 발견하는 작업을 한다. 발견된 빈발 항목은 frequent_item Table에 저장되며 전체 고객별과 개별 고객별로 생성된 빈발 항목을 저장한다.

(그림 4)는 (그림 3)에서 설계한 상품 추천 엔진의 클래스 Diagram이다.

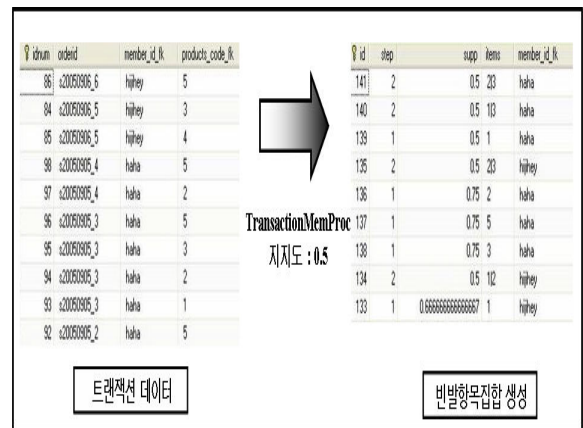


(그림 4) 상품 추천 엔진 Diagram

RecommendRng클래스에서 상품 추천 엔진이 시작되며 최소지지도와 최소신뢰도를 설정하게 된다. 또 Recommender클래스에서 추천 물을 전달 받는다. Recommender클래스는 추천 물이 이미 생성되었다는 가정하에 처리되며 추천된 상품 정

보를 상품 리스트, 추천 품 형식으로 제공해주는 기능을 한다. RuleCreator클래스는 RelationRule 클래스를 통해서 상품 추천 Rule을 생성한다. RelationRule클래스는 트랜잭션 데이터의 처리를 담당하며 연관규칙을 이용한 상품 추천 물을 실행하며 개선된 Apriori알고리즘은 이 클래스에서 한다. Transaction클래스는 트랜잭션 데이터를 디비에 추가, 제거하는 등 기본적인 정보를 처리한다. TransactionMemProc클래스는 트랜잭션 데이터를 메모리에 세팅하여 관리하며, 이를 OrderNode클래스 LinkedList에 담아서 처리하며 후보항목집합을 생성하는 작업을 처리한다. OrderNode클래스는 트랜잭션에 대한 데이터를 관리하며 트랜잭션 데이터는 주문번호, 고객 아이디, 상품 코드 리스트로 구성된다. ConstantData와 Trace는 로그정보를 관리하는 클래스이다.

(그림 5)은 (그림 4)에서 설계한 엔진으로 빈발 항목 집합을 생성한 것이다.



(그림 5) 지지도를 0.5로 설정하여 빈발 항목 집합 생성

RelationRule 모듈에서 왼쪽의 트랜잭션 데이터의 지지도를 계산하여 설정해 놓은 지지도(0.5) 이상의 데이터를 뽑아 오른쪽의 빈발 항목 집합으로 선정하여 저장하였다. RuleCreator는 RelationRule 클래스를 통해서 상품 추천 규칙을 생성하는 메인 클래스이다. 생성된 상품 추천 규칙은 recommend_rule 테이블에 저장된다.

프로그램이 수행되는 과정을 로그파일로 만들어 빈발 항목 집합과 추천 물이 생성되는 과정을



(그림 8) 로그인한 고객별 추천 상품

(그림 9)는 haha 고객이 장바구니에 물건을 넣은 후 계속 쇼핑하기를 원할 때 추천 상품 모습이다. 즉, 장바구니에 (FR-A053D)라는 냉장고가 1개 있을 경우 추천해줄 상품을 찾아 추천해준 모습이다. 장바구니에 물건이 몇 개가 있는 상품 추천해주는 과정은 같다.



(그림 9 'haha'고객에게 상품 상호 관계를 주어 추천한 모습

4. 결 론

전자상거래의 수요가 늘면서 오프라인 상점과는 차별화된 고객과의 접근방식이 필요하게 되었

다. 고객과의 효과적인 커뮤니케이션과 고객과의 장기적인 관계를 유지하기 위해 추천 시스템을 개발해 개인별 차별화된 서비스를 제공하고 멀티미디어 기능을 효과적으로 활용해야한다. 하지만 기존의 추천 시스템은 시스템을 개발하고 관리하는데 과도한 비용과 시간이 소요되는 점과 관리자가 시스템을 올바르게 이해하고 관리하는데 많은 어려움이 있다. 이를 보완하기 위해 좀 더 쉽고 이용이 간편하며 비용이 적게 드는 소규모 개인화 서비스에 대한 연구가 진행 되고 있다.

본 논문에서는 상품 추천 시스템을 설계 및 구현하기 위해 웹 마이닝 기반에서 사용자가 구매한 데이터들 사이의 연관성을 찾는 방법을 이용하였다. 데이터간의 연관성을 찾아내는 연관 알고리즘에 대한 연구 또한 다수 수행되어 실제 상용시스템에서 여러 가지 알고리즘 형태로 응용되고 있다. 본 논문에서 사용한 알고리즘은 연관규칙을 찾아주는 알고리즘 중 가장 먼저 개발된 알고리즘으로 가장 많이 쓰이고 있는 Apriori 알고리즘을 개선한 개선된 Apriori 알고리즘이다.

또 기존의 추천 시스템은 고객이 회원 가입 시 관심항목을 직접 설정해야하는 등 쇼핑몰에서 자신에게 맞는 추천을 받으려 한다면 어느 정도의 작업을 수행했어야 했다. 하지만 본 논문에서는 추천 기술 중 항목 대 항목 상관관계 기술을 사용하여 고객이 물건을 구매한 데이터만 존재한다면 별도의 작업 없이 상품 추천이 가능하다는 장점이 있다.

본 논문에서 개발한 상품 추천 시스템을 인터넷 쇼핑몰을 직접 구축하여 적용시켜 고객이 편리하게 추천을 받을 수 있는지 운영하였다. 쇼핑몰에서 직접 운영해 본 결과 상품 추천을 위해 회원 가입 시나 쇼핑하는 중간 별도의 작업을 수행하지 않기 때문에 자동화된 추천이 가능했다.

참 고 문 헌

[1] D. Amo(2000)r, The E-Business (R)evolution Living and Working in an Interconnected World, Prentice Hall
 [2] M. Eirinaki, M. Vazirgiannis(2003), "Web

Mining for Web Personalization", ACM Transactions on Internet Technology, Vol.3, No.1, pp1-27

- [3] H. Lee, W. Park and D. Park(2004), "An Efficient Algorithm for Quantitative Association Rules to Raise Reliance of Data", Lecture Notes in Computer Science, APWEB, China, pp506-512
- [4] R. Kosala, H. Blockee(2000)l, "Web Mining Research: A Survey", *SIGKDD Explorations*, Vol. 2, No. 6, pp1-15
- [5] S. Chakrabarti, B. Dom, D. Gibson, J. Kleingerg, S. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins(1999). "Mining the link structure of the world wide web," *IEEE Computer*, Vol. 32, No. 8, pp60-67
- [6] R. Cooley, B. Mobaser, and J. Srivastava(1997). "Web mining: Information and pattern discovery on the world wide web," In Proceeding of the 9th IEEE International Conference on Tools with Artificial Intelligence(ICTAI'97)
- [7] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth(1994). "From data mining to knowledge discovery: An overview,". In Proc. ACM KDD
- [8] M. A. Hearst(1999). Untangling text data mining . In Proceeding of ACL'99: the 37th Annual Meeting of the Association for Computation Linguistics

(교육학학사)

1995 한국교원대학교

컴퓨터교육과(교육학석사)

1998 ~ 현재 한국교원대학교

컴퓨터교육과

박사과정

관심분야: 컴퓨터교육, WBI

E-Mail: ssb@comedu.knue.ac.kr

박 두 순

1981 고려대학교 수학과

(이학사)

1983 충남대학교 전산학과

(이학석사)

1988 고려대학교 이과대학(전산학전공, Ph. D)

1985 ~ 현재 순천향대학교 정보기술공학부 교수

2000 ~ 현재 한국멀티미디어학회 편집위원

2002 ~ 2003 순천향대학교 공과대학 학장

2004 ~ 2005 미국 콜로라도 대학교 객원교수

관심분야: 데이터마이닝, 유비쿼터스 컴퓨팅, 병렬처리, 컴퓨터 교육

E-Mail: parkds@sch.ac.kr

김 지 혜

1991 인천교육대학교 교육학과