

# 연관 마이닝과 고객 선호도 기반의 인터넷 상품 검색 시스템 설계 및 구현

황 현 숙\*, 어 윤 양\*\*

## Design and Implementation of Product Searching System on Internet using the Association Mining and Customer's Preference

Hwang, Hyun-Suk, Eh, Youn-Yang

Most of searching systems used by shopping-mall provide too much information for user requirements or fail to provide appropriate items reflecting customer's preference.

This paper aims to design and implement the product searching systems based on customer preference which will enable efficient product selection in the internet shopping-mall. The proposed system consists of user/provider interface, searching and model agent, data management system, and model management system. Especially, we construct the searching pattern database to support fast search using association mining method. And this system includes the customer-oriented decision model which shows the highly preferred products. Input weight value per attribute and preference level should be needed to compute priority grade of preference.

---

\* 부경대학교 경영정보학과 시간강사

\*\* 부경대학교 경영학부 교수

## I. 서론

전 세계에서 최근 인터넷을 이용하는 인구는 급속도로 증가하고 있다. 특히, 인터넷 이용자 중에서 인터넷 쇼핑몰을 이용하여 상품을 구매하는 사람들도 작금에는 매우 많다. 한국에서만 개인이 운영하는 소규모 쇼핑몰을 포함하여 약 3000개 이상의 쇼핑몰 사이트가 개설되어 있는 실정이다.

KNP(<http://knp.adic.co.kr>)의 자료에 의하면 인터넷 쇼핑몰에서의 소비자들이 상품을 구매하는 이유는 "시간 절약"이라는 응답이 68.4%로 가장 많았으며, 다음이 "구매 편리"와 "가격 저렴" 등으로 조사되었다. 이는 인터넷 쇼핑몰에서 빠른 검색 정보에 관련된 요소가 쇼핑몰 사이트의 중요한 경쟁 요소임을 나타내고 있다.

인터넷 쇼핑몰에서 효과적인 상품 정보를 제공하기 위한 중요한 요소로 다음과 같은 것을 생각해 볼 수 있다.

첫째, 고객의 선호도에 적합한 상품을 우선적으로 제시하여 상품 선정에 대한 의사결정을 지원하는 동적인 정보 서비스를 제공하는 것이다. 인터넷 쇼핑몰에서 제공되는 상품 정보는 데이터베이스 검색에 의해 일반적으로 고객에게 판매하려는 전체 상품의 목록만을 제시하는 경우가 대부분이다. 이 때문에 상품 목록이 많을 경우 고객은 적합한 상품을 선택하기 어렵고 때로는 쇼핑몰에서의 상품 구매를 포기하는 경우도 발생할 수 있다. 그러므로 상품 선정을 쉽게 할 수 있는 방법이 필요하다.

둘째, 고객에게 시스템적으로 빠르고 효과적인 정보를 제공하는 것이다. 고객에게 제공되는 정보는 데이터베이스에서 검색되기 때문에 데이터베이스의 크기와 통신 속도는 정보 전달에 가장 중요한 요소이다. 이에 따라 효과적으로 데이터베이스를 구성하는 것은 정보 전달 속도 측면에서 중요하다.

현재까지 인터넷 쇼핑몰에서 효과적이고 빠

르게 상품 정보를 제공하고자 하는 기존의 연구는 있었으나 이들의 연구는 단순히 정보 검색의 속도를 향상시키거나 콘텐츠의 내용 구성에 대한 연구가 대부분이었다[Guttman et al., 1998; 이재규 외, 1999].

이러한 측면에서 인터넷 쇼핑몰에서 고객에게 필요한 상품 정보를 효과적으로 전달하는 방법에 대한 연구가 현 시점에서 매우 중요하고 시급한 과제로 인식되고 있다. 본 논문에서는 이러한 배경에서 쇼핑몰에서 효과적인 상품 정보를 제공하기 위한 연구를 수행하였다. 쇼핑몰에서 고객이 선호하는 상품의 속성에 따라 키워드 검색을 수행할 때 빠른 상품 검색 결과를 제공할 수 있도록 하고, 상품 선정에 대한 의사결정을 지원할 수 있는 상품 검색 시스템을 개발하고자 한다. 빠른 상품 검색을 제공하기 위해서는 연관 데이터 마이닝 방법을 사용하고 고객의 상품 선정을 도와주기 위해서 다속성 의사결정 모델을 이용한다.

본 논문의 구성은 제 2장에서 연관 마이닝 방법과 인터넷 의사결정지원 시스템에 대해서 설명하고 제 3장에서는 의사결정 기능을 포함하고 있는 상품 검색 시스템을 설계한다. 그리고 제 4장에서는 연관 마이닝을 이용한 상품 검색 시스템을 구현한다. 제 5장에서는 효율적인 검색 유형 데이터베이스를 구성하기 위한 성능을 분석하였다. 마지막으로 제 6장에서는 연구의 결과와 향후 연구과제를 제시한다.

## II. 선행 연구

### 2.1 연관 마이닝 방법

데이터마이닝은 데이터 집합에 숨겨져 있으면서 논리적인 구성이 가능하고 유용할 것으로 생각되는 패턴을 발견하기 위한 특별한 과정으로 정의된다[Adriaans et al., 1996; Berson et al., 1997; Fayyad et al., 1995]. 이러한 데이터

마이닝은 특정 분야에만 활용되는 것은 아니다. 마케팅, 소매업, 금융, 생산, 건강과 의학, 에너지와 공익 설비 등 모든 분야에서 활용될 수 있다. 예를 들어 데이터베이스 마케팅 분야에서는 고객유치, 특별한 판촉에 대한 정보를 제공받을 수 있다. 시장바구니 분석에서는 상품간의 연관성을 발견하여 상품 배치와 판매촉진 전략 등에 대한 정보를 제공받을 수 있다. 은행 및 금융 분야에서는 고객 신용카드의 잠재적 사기 유형에 대한 정보를 제공받을 수 있다. 그리고 보험금 청구 금액을 분석하여 사기성이 있는 행동에 대한 정보를 제공받을 수 있다. 건강 및 의학에서는 환자 행동을 특징지어 병원 방문을 예측할 수 있는 정보를 제공받을 수 있다.

데이터 마이닝 방법 중에서 특히, 연관 방법은 데이터베이스의 트랜잭션에서 항목간에 발생하는 규칙을 표현하는 것으로 1993년에 처음 소개되었다[Agrawal et al., 1993]. 연관은 어떤 사건이 발생할 때 그 다음 사건의 관련성을 의미하는 것으로  $X \Rightarrow Y$  규칙의 형태로 표현한다.  $X \Rightarrow Y$ 의 규칙은 데이터베이스의 트랜잭션 중  $X$ 라는 항목 집합을 포함하는 트랜잭션은 항목 집합  $Y$ 도 함께 포함하는 경향이 있음을 의미한다.

연관성을 발견하기 위한 초기의 연구로 Agrawal [1993]는 장바구니 데이터를 대상으로 고객이 구매한 상품간에 연관성이 있는 집합을 발견하는 AIS(Association Item Sets) 알고리즘을 제안하였다. 연관 알고리즘은 대부분 두 단계로 구성되어 있는데, 첫 번째 단계는 최소 지지도 이상을 갖는 빈발 항목집합을 발견하는 단계이고 두 번째 단계는 발견된 빈발 항목집합의 모든 부분 집합을 생성하여 최소의 신뢰도 이상인 규칙을 발견하는 단계이다.

연관 규칙은 지지도(support)와 신뢰도(confidence)로 타당성이 판단되어진다. 지지도는 전체 트랜잭션에 대해 트랜잭션 항목 집합이 차지하는 비율을 의미하고 신뢰도는 조건부 트랜잭션 항목 집합에 대해 규칙에 포함되는 모든

항목 집합이 차지하는 비율을 의미한다[Agrawal et al., 1993; Srikant et al., 1993; Megiddo & Srikant, 1998].

## 2.2 인터넷 의사결정지원시스템

의사결정지원시스템(Decision Support System; DSS)은 의사결정자가 비구조적인 문제를 해결할 때 데이터와 모델을 이용하여 의사결정을 하게 도와주는 대화식의 컴퓨터 기반 시스템이다 [Morton, 1971; Sprague, 1980; Turban, 1995; Sauter, 1997; Gio, 2000].

의사결정지원시스템은 개인 의사결정지원시스템과 그룹 의사결정지원시스템 형태로 발전되어 왔다. 최근 인터넷이 활성화됨으로써 의사결정지원시스템은 시간 및 공간을 초월하여 의사결정을 도와주는 인터넷 의사결정지원시스템으로 발전되고 있다.

개인 및 그룹의 의사결정을 지원하는 전통적인 의사결정지원시스템은 한정된 데이터, 모델, 사용자와의 인터페이스 모듈을 이용하여 한 명 또는 여러 명의 문제를 해결하는 시스템이다. 여기에서 발전된 인터넷 의사결정지원시스템은 전통적인 의사결정지원시스템과 별도의 개념이 아니라 유사성을 가지면서 인터넷이라고 하는 통신망에서 지능형 에이전트에 기초한 보다 발전된 형태의 시스템으로 정의된다[Bhargava et al., 1995, 1998; 이진창 외, 1997; Chou, 1998, 황현숙 & 어운양, 1999].

인터넷 의사결정지원시스템은 다음과 같은 특성을 가지고 있다.

첫째, 복수개의 의사결정지원시스템은 인터넷이라는 개방형 정보통신 환경에서 서로 연결되어 있다. 그러므로 인터넷 의사결정지원시스템은 기존의 개인 의사결정지원시스템이나 그룹 의사결정지원시스템의 특성을 포함한다.

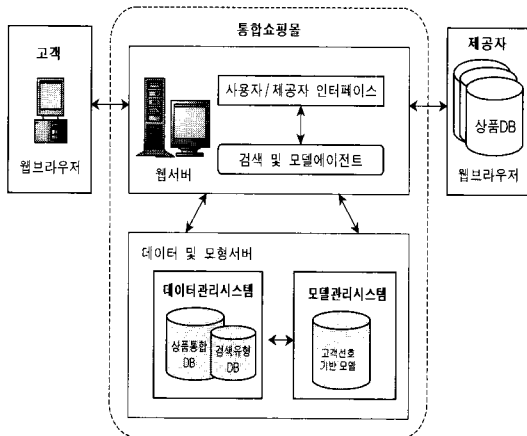
둘째, 인터넷 의사결정지원시스템은 지능형 에이전트를 구성요소로 가지고 있어야 한다. 웹

에서 지능형 에이전트는 정보 검색, 모델 구축 그리고 결과해석 등의 여러 활동을 능동적으로 수행하여 의사결정자의 문제를 해결할 수 있도록 도와준다.

### Ⅲ. 상품 검색 시스템 설계

#### 3.1 상품 검색 시스템 구조

쇼핑몰 상품 검색에서 고객 선호도를 기반으로 의사결정을 지원하는 상품 검색 시스템 통합 모듈은 <그림 1>과 같이 구성하였다.



<그림 1> 상품 검색 시스템 구성

이 시스템은 상품 검색을 수행하는 고객, 상품을 제공하는 제공자, 사용자/제공자 인터페이스, 검색 및 모델 에이전트, 데이터관리 시스템, 모델관리 시스템 등으로 구성되어 있으며, 이는 전통적인 의사결정지원시스템의 기본 프레임에서 제시하는 구성 요소를 모두 포함하고 있다.

사용자 인터페이스는 고객과 통합쇼핑몰간의 인터페이스로서 검색 키워드를 입력하고 검색 정보를 제공하도록 연결시키는 부분이다. 제공자 인터페이스는 상품 데이터를 보관하고 있는 제공자의 상품 데이터와 통합쇼핑몰을 연결시키는 부분이다.

데이터관리 시스템은 빠른 상품 검색을 지원하기 위해 두 종류의 데이터베이스를 운영한다. 첫째는 상품 전체의 데이터를 가지고 있는 상품 통합 데이터베이스이고 두 번째는 연관 마이닝 방법을 이용하여 전체 상품 데이터베이스에서 자주 검색하는 유형의 데이터를 발견하여 검색 유형 데이터베이스를 생성하여 운영한다. 이러한 검색 유형 데이터베이스를 운영함으로써 고객에게 빠른 상품 검색을 제공할 수 있게 된다.

모델관리시스템은 데이터베이스에 저장되어 있는 데이터로 모델을 구축하여 사용자에게 동적인 정보를 제공하게 된다.

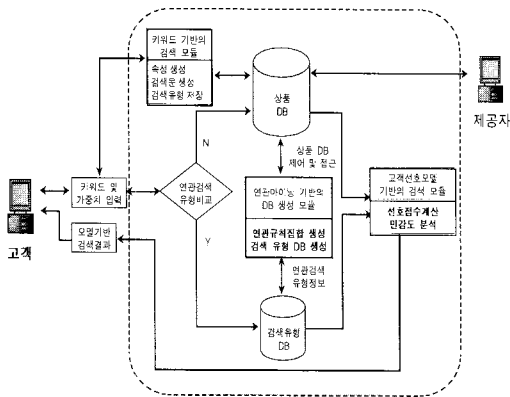
검색 및 모델 에이전트는 고객이 쇼핑몰에서 검색을 수행할 때 적합한 상품을 제공받을 수 있도록 도와주는 역할을 한다. 검색 에이전트는 고객이 선호하는 속성을 선택할 수 있도록 상품 테이블에 있는 필드의 목록을 제공한다. 모델 에이전트는 모델에 대한 전문적인 지식이 부족한 고객을 위하여 모델에 필요한 입력 및 출력 정보를 알려주는 역할을 한다.

#### 3.2 시스템 구현 모듈 구성

고객에게 빠른 상품 검색과 상품 선정에 대한 의사결정을 지원하는 시스템을 구현하기 위한 내부 구성 모듈은 <그림 2>와 같다.

키워드 기반의 검색 모듈은 고객이 선호하는 속성 및 검색문 생성과 검색한 상품 유형을 저장하는 기능을 가지고 있다. 연관 마이닝 기반의 데이터베이스 생성 모듈은 수집한 검색 유형 데이터를 가지고 빠른 검색 결과를 제공하기 위해 주기별로 검색 유형 데이터베이스를 생성한다. 그리고 고객 선호 모델 기반의 검색 모듈은 고객이 검색한 상품 속성의 가중치 및 선호 기준을 입력받아 선호점수를 계산하여 점수가 높은 상품을 우선적으로 제공하는 기능과 여러 번 검색한 결과를 분석하는 민감도 분석 기능

을 가지고 있다.



<그림 2> 시스템 구현 모듈 구성도

<그림 3>은 상품 데이터를 대상으로 시스템 구축에 필요한 데이터베이스의 구조이다. 상품 데이터 테이블은 상품 데이터를 보관하고 있고 검색 유형 데이터 테이블에는 고객이 자주 검색

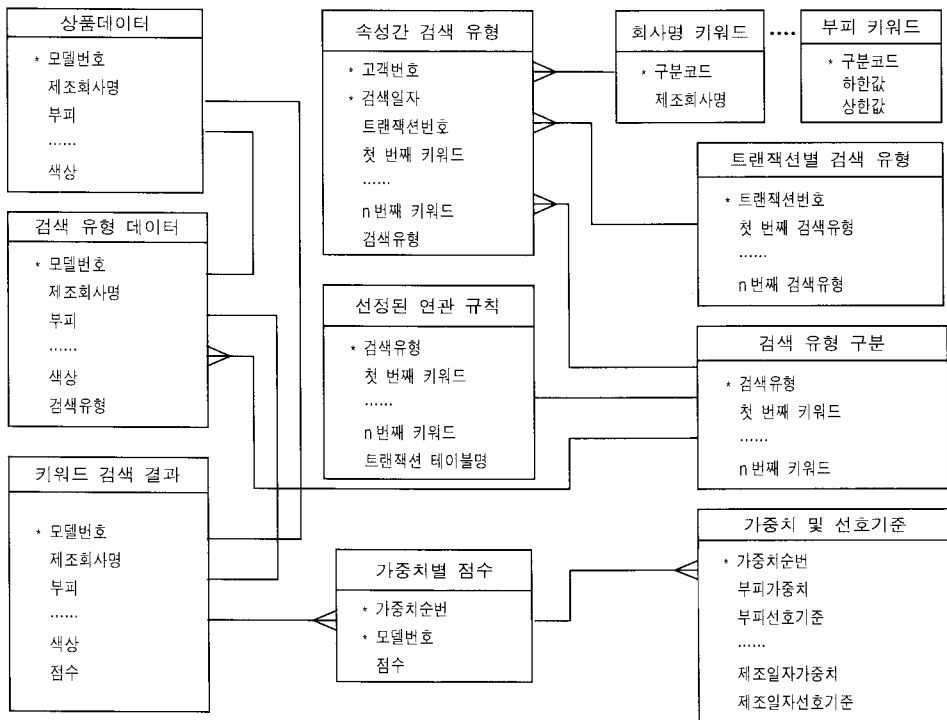
하는 검색 유형 데이터를 보관하고 있다. 이는 연관 마이닝 방법을 이용하여 데이터베이스를 구축한다. 고객이 검색한 데이터는 키워드 검색 결과 테이블에 저장해 두어 검색한 상품에 대해 고객 선호도에 따라 다수의 검색을 수행할 수 있도록 한다.

### 3.2.1 키워드 기반 검색 모듈

키워드 기반 검색 모듈은 입력한 상품의 키워드에 대해 축약된 정보를 제공하는 모듈이다. 이 모듈에서는 속성 및 검색문을 자동으로 생성하여 고객이 선호하는 상품 속성에 따라 키워드 검색을 수행할 수 있도록 한다.

#### (1) 속성 및 검색문 생성

속성 생성에서는 고객이 검색하려는 상품을 선택하면 상품 데이터 테이블에 있는 속성의 목



<그림 3> 데이터베이스 구조

록을 제공하게 된다. 즉, 고객이 선호하는 속성을 직접 선택하여 검색할 수 있다.

검색문 생성은 고객이 선택한 속성에 따라 키워드를 자동적으로 생성하여 검색문을 제공한다. 키워드는 속성별로 고객이 선택할 수 있는 데이터 또는 데이터 범위를 의미한다. 고객이 선택한 속성에 따라 키워드를 생성하기 위해서는 속성별로 키워드에 대한 테이블이 필요하다.

(2) 검색 유형 저장

검색 유형 저장에서는 고객에게 빠른 상품 검색을 제공하기 위해 고객이 검색한 키워드를 입력받아 두 가지의 형태로 데이터를 저장한다. 첫 번째는 단일 상품에서 속성간의 검색 유형 데이터를 저장하고 두 번째는 단일 상품에서 트랜잭션별 검색 유형 데이터를 저장한다.

<표 1>은 고객이 검색한 키워드 검색 유형을 저장하기 위해 필요한 속성간 검색 유형을 저장하기 위한 테이블이다. 고객은 쇼핑몰에서 단일 상품에 대해 다수의 검색을 수행할 수 있으며, 트랜잭션 번호(T\_id)는 고객이 쇼핑몰에 로그인 해서 키워드 상품 검색을 수행할 때마다 누적된다.

<표 1> 속성간 검색 유형 테이블

C_id	Date	T_id	Item <sub>1</sub>	Item <sub>2</sub>	...	Item <sub>n</sub>	S_kind
고객 번호	검색 일자	트랜잭션 번호	첫 번째 속성의 키워드	두 번째 속성의 키워드	...	n 번째 속성의 키워드	검색 유형

검색 유형(S\_kind)은 트랜잭션별 연관 집합을 발견하기 위해서 고객이 검색한 유형을 저장하는 필드이다. 검색 유형 필드에 고객이 검색한 유형을 표현하기 위해서는 모든 속성에서 가능한 검색 유형이 있는 검색 유형 구분 테이블이 필요하다.

속성간 검색 유형은 고객이 수행한 키워드 검색의 전체 트랜잭션을 대상으로 하기 때문에

최소한 두 개 이상의 속성에서 연관 관계가 이루어진다. 이는 고객이 단일 속성만을 선택할 경우에는 연관 규칙을 반영하지 못하는 단점이 발생한다. 따라서 트랜잭션별 검색 유형 데이터를 사용하여 연관 규칙을 적용하여 고객이 단일 항목을 선택할 때도 연관 규칙이 반영될 수 있도록 한다. 고객이 인터넷 쇼핑몰에 들어와서 단일 상품에 대해 자신이 원하는 여러 항목에 대한 검색 결과를 제공받는 과정을 한 번의 트랜잭션으로 정의한다. 트랜잭션별 검색 유형 데이터를 생성하기 위해 <표 1>의 속성간 검색 유형 테이블에서 (C\_id, Date) 두 항목의 동일한 값을 근간으로 트랜잭션 번호(T\_id)를 부여하여 <표 2>의 트랜잭션별 검색 유형 테이블을 생성한다.

<표 2> 트랜잭션별 검색 유형 테이블

T_id	S_kind <sub>1</sub>	S_kind <sub>2</sub>	...	S_kind <sub>n</sub>
트랜잭션 번호	첫 번째 트랜잭션의 검색유형	두 번째 트랜잭션의 검색유형	...	n 번째 트랜잭션의 검색유형

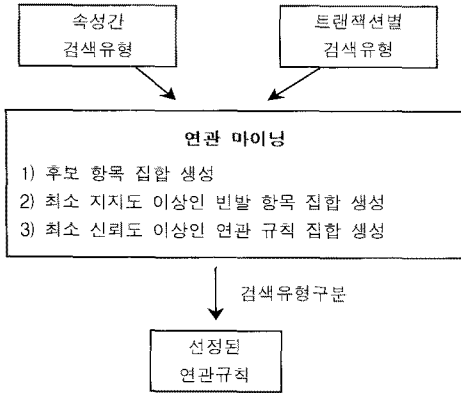
3.2.2 데이터베이스 생성 모듈

데이터베이스 생성 모듈은 빠른 상품 검색을 제공하기 위해 연관 마이닝 방법을 이용하여 검색 유형 데이터 베이스를 구축한다.

(1) 연관 규칙 집합 생성

연관 규칙 집합 생성에서는 검색 유형 데이터베이스를 생성하기 위해 고객이 검색한 검색 유형 데이터를 이용하여 최소 지지도와 신뢰도 이상을 가지는 연관 규칙 집합을 선정한다. 우선, 검색 유형 데이터베이스를 구성하기 위해서 속성간 및 트랜잭션별 검색 유형 데이터에 대해 연관 마이닝을 적용하여 최소 지지도와 신뢰도 이상을 가지는 연관 규칙 집합을 생성하여야 한다. <그림 4>는 두 가지 형태의 검색 유형 데이터에서 연관 마이닝 방법을 적용하여 선

정된 연관 규칙 집합을 생성하는 과정이다.



<그림 4> 선정된 연관 규칙 집합 생성

생성된 연관 규칙을 저장하는 테이블 구조는 <표 3>과 같다. 이는 빈발 항목 집합에서 최소 지지도와 신뢰도 이상을 가지는 트랜잭션 유형을 저장하는 테이블이다. 이 테이블에는 연관성이 있는 속성의 키워드를 보관하고 있기 때문에 이에 해당하는 실제 데이터를 전체 상품 데이터베이스에서 가져와야 한다. 트랜잭션 테이블명은 전체 상품 데이터베이스에서 검색 유형에 해당하는 데이터를 보관하고 있는 테이블 이름이다.

<표 3> 선정된 연관 규칙 테이블

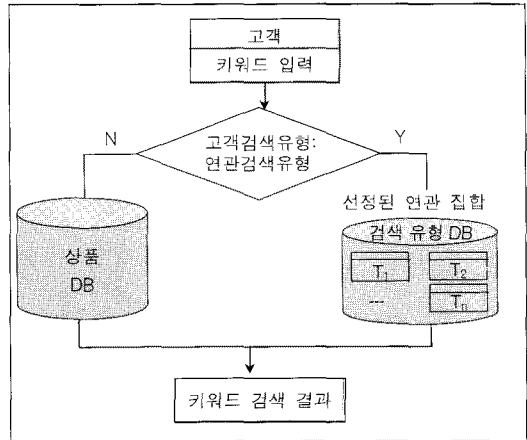
S_no	S_kind	Item <sub>1</sub>	Item <sub>2</sub>	...	Item <sub>n</sub>	T_name
순번	검색 유형	첫 번째 속성의 키워드	두 번째 속성의 키워드	...	n 번째 속성의 키워드	트랜잭션 테이블명

(2) 검색 유형 데이터베이스 생성

검색 유형 데이터베이스 생성에서는 선정된 연관 규칙 집합에 해당하는 데이터를 전체 상품 데이터베이스에서 가져와 검색 유형별로 테이블에 저장시킨다. 이는 연관 규칙 집합으로 선정된 검색 유형별로 해당하는 검색 결과 데이터를 저장하기 위해 필요한 테이블이다.

3.2.3 검색 유형 비교

제기한 시스템은 빠른 상품 검색을 위해서 별도의 검색 유형 데이터베이스를 운영하고 있다. 이 때의 검색 과정은 다음과 같다. 고객이 입력한 키워드 검색 유형과 검색 유형 데이터베이스에 있는 연관 검색 유형과 동일하면 검색 유형 데이터베이스에서 검색 결과를 제공하고, 그렇지 않으면 통합 데이터베이스에서 검색 결과를 제공한다. <그림 5>는 이러한 상품 검색 과정을 나타내고 있다. 검색 유형 데이터베이스는 선정된 n개의 연관 규칙 집합에 대해 n개의 검색 유형 데이터 테이블에 대한 정보를 가지고 있다. 고객이 키워드로 검색한 데이터는 키워드 검색 결과 테이블에 저장되는데, 이는 키워드 검색 결과에 대해 고객에게 알맞은 상품을 우선적으로 제시하기 위해 필요하다.



<그림 5> 상품 검색 과정

3.2.4 고객 선호 모델 기반의 검색 모듈

모델 기반 검색 모듈에서는 의사결정을 지원하는 모델을 가지고 있어 상품 선택을 위한 의사결정을 지원한다.

(1) 선호 점수 계산

상품의 여러 가지 속성을 고려하여 의사결정

을 지원하는 다속성 의사결정 모델에서 고객에게 적합한 의사결정을 지원하기 위해서는 여러 가지 모델들이 있는데, 본 연구에서는 점수 모델의 단순 가중치 방법을 사용한다[황현숙 & 어윤양, 1999]. <그림 6>은 고객의 선호도에 따른 상품의 점수를 계산하는 알고리즘이다.

```

Algorithm Jumsu_Computation
JUMSU(i): 상품별 선호 점수,
D(i,j): 레코드의 항목 속성값
ND(i,j): 각 속성의 정규치
PRE(j): 각 속성의 선호기준
      (1: 이윤속성, 0: 비용속성)
Max[Di(i,j)]: 각 속성의 최대값,
Min[Di(i,j)]: 각 속성의 최소값
W(j): 각 항목의 가중치
      (0 ≤ W(j) ≤ 1, ∑j=1m W(j) = 1)
for(i=1, n; i++) do begin /* i번째 레코드
  for(j=1, m; j++) do begin
    /* i번째 레코드의 j번째 항목
    if (PRE(j) = 1) then
      ND(i,j) = D(i,j) / Max[Di(i,j)]
    else
      ND(i,j) = Min[Di(i,j)] / D(i,j)
    endif
    JUMSU(i) = JUMSU(i) + W(j) * ND(i,j)
  end
end
    
```

<그림 6> 고객 선호도에 따른 점수 계산 알고리즘

고객의 선호도를 고려하기 위해 고객이 선호하는 가중치와 선호기준이 필요하다. 고객이 입력한 가중치의 합은 1이어야 하고 속성의 선호기준은 속성 데이터의 값이 높을수록 높은 선호인 이윤 속성과 속성 데이터의 값이 낮을수록 낮은 선호를 가지는 비용 속성으로 구분한다. 데이터의 정규화는 선호 속성에 따른 선형 변환 방법을 사용하여 계산한다. 그런 후 선호 점수는 고객이 입력한 속성별 가중치와 선호기준에 따라 정규화한 데이터를 곱해서 구한다.

이러한 모델 기반 검색을 위해 필요한 데이

블은 다음과 같다. 고객이 키워드 입력으로 검색한 데이터를 저장하는 키워드 검색 결과 테이블이 필요하다.

여기서 점수(Jumsu) 필드의 값은 고객이 가중치와 선호기준을 입력하였을 때 계산된다. 가중치 및 선호 기준 테이블은 점수 계산에 사용될 고객이 입력한 선호 가중치와 선호기준을 저장하기 위해 필요하다. 가중치별 점수 테이블은 민감도 분석에서 사용하기 위해 선호 점수 계산에서 나온 다수의 검색에 대한 점수를 저장하기 위한 필요하다.

#### (2) 민감도 분석

고객은 선호하는 속성에 대해 가중치의 반영 비율을 변화하면서 모델 기반 검색을 수행할 수 있다. 이러한 검색 결과를 분석하기 위해 민감도 분석을 수행한다. 민감도 분석은 고객이 입력한 가중치 변화에 따라 상품 순위의 변화를 분석한다. 따라서 고객은 다양한 가중치를 제시함으로써 적합한 상품을 선정하는 데 도움을 받게 된다. 민감도 분석 업무에서는 가중치 및 선호기준 테이블과 가중치별 점수 테이블을 입력받아 고객에게 점수 변화에 따른 상품의 정보를 제공하게 된다.

## IV. 연관 마이닝 통합 시스템 구현

### 4.1 구현환경

본 논문에서 설계한 상품 검색 시스템은 Windows NT 운영체제에서 웹서버는 마이크로 소프트웨어사의 IIS4.0와 썬 마이크로시스템의 Tomcat 3.2를 사용하고 데이터베이스는 MSSQL Server 2000을 사용하여 개발하였다. 개발 툴로 Java 언어를 사용하고 데이터베이스와의 연동을 위해 JDBC 드라이버(aveConnect JDBC2.3)를 사용하였다[Art, 1997; Ian et al., 1999]. 그리고 웹서버와 데이터베이스 연결 및 동적 데이터를 처

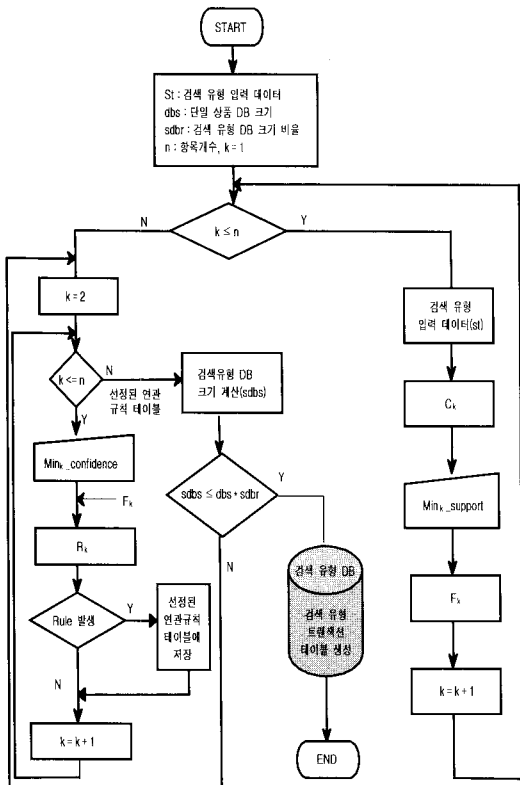


리하기 위해 Tomcat 엔진을 통한 JSP(Java Server Page)를 사용하였다[Pekowsky, 2000].

#### 4.2 연관 마이닝에 의한 검색 유형 DB 생성

검색 유형 데이터베이스를 구축하기 위해서 우선, 상품의 키워드 검색 화면에서 고객이 검색한 검색 유형을 수집하여야 한다. 이 때 수집한 검색 유형 데이터는 3.2.1절에서 제시한 속성간 및 트랜잭션별 검색 유형 테이블에 저장된다. 연관 알고리즘 구현을 위해 Java 언어를 사용하였고 데이터베이스와 연결하기 위해 Atinav사의 JDBC 드라이버를 사용하였다.

이터베이스를 생성하는 흐름도이다. 연관 규칙 집합을 생성할 때 최소 지지도(Min<sub>k</sub>\_support) 및 신뢰도(Min<sub>k</sub>\_confidence) 이상을 가진 집합은 검색 유형 데이터베이스 생성에 사용되기 때문에 규칙이 발생되면 선정된 연관 규칙 테이블에 저장해 둔다. 이를 n단계까지 수행한 후 생성된 연관 규칙 테이블에서 검색 유형 데이터베이스의 크기를 계산한다. 계산한 검색 유형 데이터베이스의 크기가 설정한 단일 상품 데이터베이스 비율의 크기보다 적거나 같으면 검색 유형 데이터 테이블을 생성하고 그렇지 않을 경우에는 k=2부터 최소 신뢰도를 다시 설정하여 연관 규칙 테이블을 발견한다.



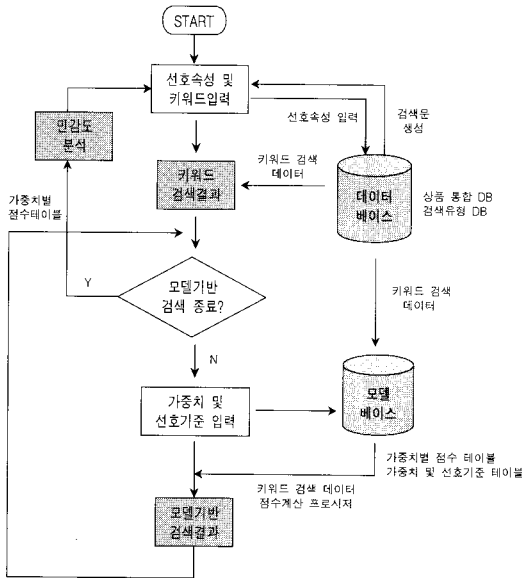
<그림 7> 검색 유형 데이터베이스 생성 흐름도

<그림 7>은 후보(C<sub>k</sub>) 및 빈발(F<sub>k</sub>) 항목 집합과 연관 규칙 집합(R<sub>k</sub>)을 발견하여 검색 유형 데

#### 4.3 고객 선호도 기반의 상품 검색

고객이 쇼핑몰에서 상품을 구매할 경우 우선, 상품의 정보를 알기 위해 상품 검색을 수행한다. <그림 8>은 상품 검색 시스템에서 고객에게 적합한 상품을 우선적으로 보여주어 상품 구매에 대한 의사결정을 지원하는 흐름도이다. 고객은 검색하려는 상품 속성별로 키워드를 입력하여 데이터베이스에서 키워드 검색 결과를 제공받는다. 제시한 시스템은 빠른 상품 검색을 제공하기 위해 별도의 검색 유형 데이터베이스를 운영하고 있다.

데이터베이스에서 추출된 키워드 검색 데이터는 모델 기반 검색을 위해 모델베이스의 키워드 검색 테이블에 저장해 둔다. 고객은 속성별로 상대적인 가중치와 선호기준을 입력하여 모델 기반 검색을 제공받게 된다. 고객이 입력한 정보를 바탕으로 모델베이스에서 상품별로 점수를 계산하여 점수가 높은 순서로 상품 검색 결과를 제공함으로써 고객의 구매 의사결정을 지원할 수 있게 된다. 그리고 고객은 키워드 검색 데이터에 대해 속성의 가중치를 여러번 변화해서 상품의 우선 순위를 제공받는 민감도 분석을 수행할 수 있다.



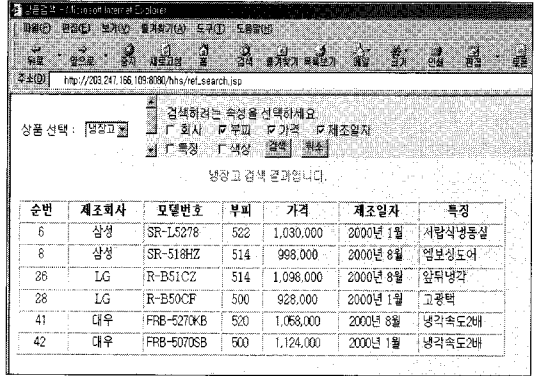
<그림 8> 상품 검색 의사결정 지원 흐름도

### 4.3.1 키워드 검색

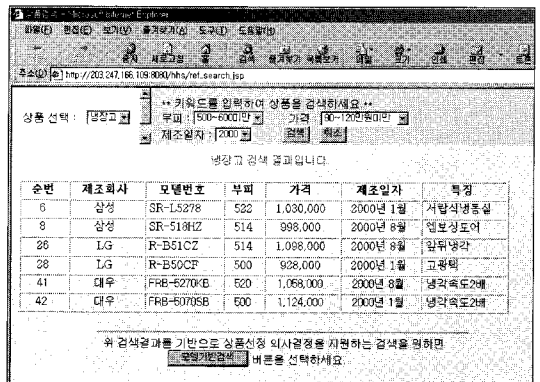
쇼핑몰에서 상품을 검색할 때 고객마다 선호하는 속성이 다를 것이다. 이러한 점을 고려하여 고객이 상품을 선택하면 검색 에이전트는 상품 데이터 테이블에 있는 속성의 목록을 제공한다. <그림 9>는 고객이 냉장고를 선택하였을 때 검색하려는 속성을 선택하는 화면이다. 만약 제공자가 상품을 제공할 때 해당 속성값을 입력하지 않았을 경우에는 해당 속성의 검색 결과는 공백으로 처리된다. 속성값이 공백인 상품 모델은 고객 선호 모델을 이용하여 상품 모델의 우선 순위를 정할 때 불리한 결과를 초래할 수 있다.

<그림 10>은 고객이 부피, 가격, 제조일자에 해당하는 키워드를 입력하여 상품을 검색하는 화면이다. 고객이 상품 키워드를 입력하여 상품을 검색하면 고객이 입력한 상품 키워드 검색 유형과 검색 유형 데이터베이스에 있는 검색 유형 데이터 테이블에 있는 검색 유형과 비교한다. 그런 후 해당 유형이 존재하면 검색 유형 데이터베이스에서 검색 결과가 추출되고 그렇지

않으면 통합 데이터베이스에서 검색 결과가 추출된다.



<그림 9> 속성 선택 화면



<그림 10> 키워드 입력에 의한 상품 검색

### 4.3.2 모델 기반 검색

<그림 10>의 키워드 검색 결과에서 모델기반검색 버튼을 선택하면 <그림 11>과 같이 모델 기반 검색에 필요한 입력값과 그에 대한 결과 화면이 표시된다. 이 화면에서 고객이 속성별 가중치의 반영 비율과 선호기준을 입력하면 모델 베이스에서 점수를 계산하여 높은 점수를 가진 상품을 우선적으로 보여주게 된다.

<그림 11>은 부피 0.5, 가격 0.4, 제조일자 0.1의 가중치를 입력한 경우로 가중치가 고려되어 28번 레코드가 고객의 선호도에 가장 적합

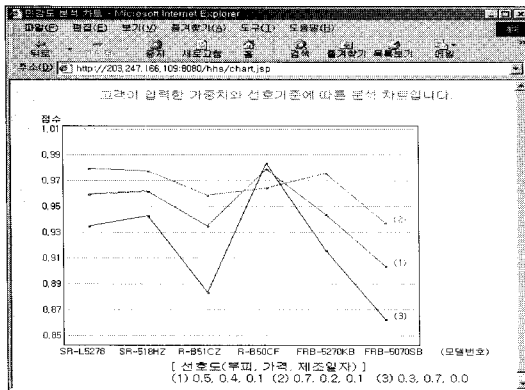
한 상품임을 나타내고 있다.

순번	제조사	모델번호	부피	가격	제조일자	특징	선호점수
28	LG	R-B50CF	500	928,000	2000년 1월	고공력	97.83
8	삼성	SR-518HZ	514	998,000	2000년 8월	정보상도어	96.43
6	삼성	SR-L5276	522	1,030,000	2000년 1월	서랍식냉동실	95.98
41	대우	FRB-5270KB	520	1,058,000	2000년 8월	냉각속도2배	94.89
36	LG	R-B51CZ	514	1,098,000	2000년 8월	왕취냉각	93.04
42	대우	FRB-5070SB	500	1,124,000	2000년 1월	냉각속도2배	90.86

<그림 11> 모델 기반 검색 결과

### 4.3.3 고객 선호도에 따른 민감도 분석

고객이 입력한 속성별 가중치 변화에 따라 상품의 순위를 표시해 주는 민감도 분석을 수행하였다. 민감도 분석은 각 속성에 대한 가중치를 다르게 할 경우 어떤 상품이 가장 많은 영향을 받는지를 판별하게 해 준다. 따라서 고객은 다양한 가중치를 제시함으로써 고객에게 가장 적합한 상품을 선정 받을 수 있게 된다.



<그림 12> 고객 선호도별 민감도 분석

<그림 12>는 반영 비율을 달리한 3가지(① 부피: 0.5, 가격: 0.4, 제조일자: 0.1, ② 부피: 0.7, 가격: 0.2, 제조일자: 0.1, ③ 부피: 0.3, 가격: 0.7,

제조일자: 0.0) 경우에 대해 상품별 선호 순위를 나타낸 것이다. 분석 결과의 특징은 다음과 같다.

첫째, 첫 번째와 세 번째 가중치의 경우는 상품별로 동일한 순위를 나타낸다. 첫 번째의 경우는 부피의 가중치가 약간 높지만 가격의 가중치가 융합되어 나온 결과이고 세 번째는 가격의 가중치가 높게 설정되었기 때문에 가격이 낮은 순으로 표시된 것이다.

둘째, 8번과 28번 상품이 높은 우선 순위이며, 42번 상품은 낮은 우선 순위를 나타낸다. 결국, 부피와 가격을 우선적으로 고려할 경우에 28번과 8번 상품이 선정될 수 있고, 부피를 우선적으로 고려할 경우에는 6번과 8번 상품이 선정될 수 있다. 이러한 결과를 이용하여 고객은 자신의 다양한 요구에 적합한 상품에 대한 의사결정을 지원 받게 된다.

## V. 성능 분석

제안한 상품 검색 시스템에서는 고객에게 빠른 검색 정보를 제공하기 위해 연관 마이닝 방법을 사용하여 자주 검색하는 상품들을 별도의 검색 유형 데이터베이스에 저장해 두었다. 따라서 검색 유형 데이터베이스의 효과적인 구성을 분석하기 위해 최소 지지도 및 신뢰도에 따라 생성되는 데이터베이스의 크기를 비교하였다. 그리고 검색 유형 데이터베이스를 구성한 후, 고객이 입력한 정보가 검색 유형 데이터베이스에서 검색 결과가 제공될 적중률을 측정하였다. 또한, 고객에게 검색 결과를 응답해 주는 검색 시간을 측정하였다. 실험에 사용한 데이터는 냉장고 데이터를 대상으로 하였다. 냉장고에 관한 전체 상품 데이터 테이블은 500개의 레코드로 구성하였고 이때, 레코드의 크기는 92byte이다.

### 5.1 검색 유형 데이터베이스 크기 비교

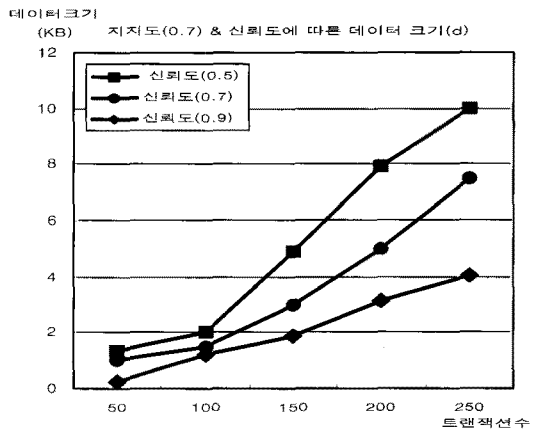
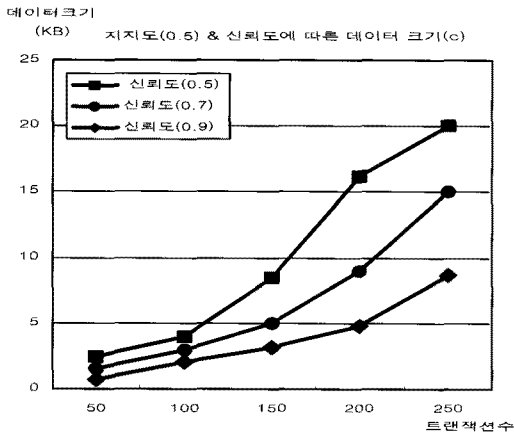
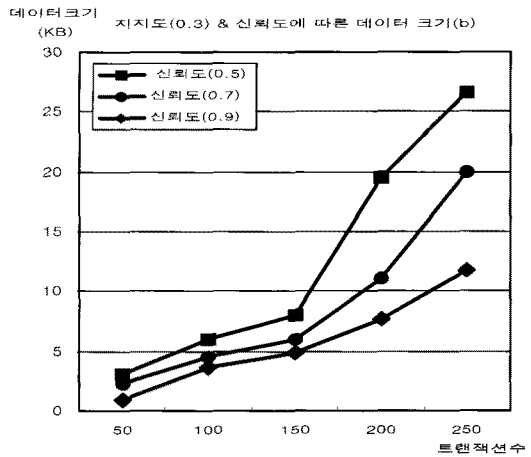
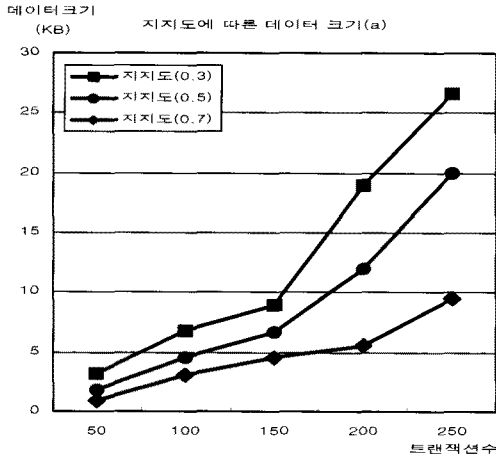
고객의 키워드 검색에서 50부터 시작하여 50

씩 증가하여 250개의 트랜잭션을 수집하여 실험에 사용하였다. 수집한 데이터는 연관 마이닝 방법을 사용하여 검색 유형 데이터베이스를 각각 구축하였다. 그런 후 지지도와 신뢰도를 변화하면서 데이터 크기를 계산하였다. 데이터 크기는 생성되는 검색 유형 데이터 테이블의 수와 각 테이블당 레코드 수, 레코드의 전체 크기를 모두 곱해서 계산하였다.

<그림 13>은 지지도와 신뢰도에 따른 검색 유형 데이터베이스의 크기를 비교한 것이다. (a)는 트랜잭션의 회수별로 지지도에 따른 데이터의 크기를 나타낸다. 각 트랜잭션 회수별로 지

지도가 높을수록 데이터 크기가 줄어들음을 알 수 있다. (b),(c),(d)는 지지도별로 신뢰도를 0.5, 0.7, 0.9로 변화함으로써 데이터 크기가 어떻게 되는지를 실험하였다. 결국, 지지도에 따라 신뢰도가 높아짐에 따라 데이터 크기가 줄어들음을 알 수 있다.

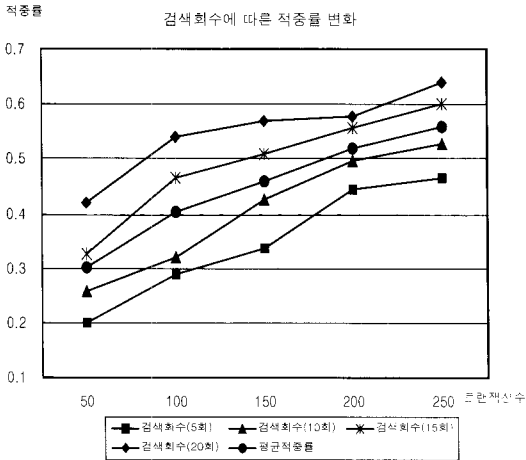
검색 유형 데이터베이스는 지지도와 신뢰도에 따라 연관 집합이 결정된다. 지지도와 신뢰도가 낮을수록 연관 집합이 많아져 상대적으로 검색 유형 데이터베이스의 크기는 커지며, 지지도와 신뢰도가 높을수록 연관 집합이 적기 때문에 검색 유형 데이터베이스의 크기는 적어진다.



<그림 13> 지지도 및 신뢰도에 따른 데이터 크기

## 5.2 검색 유형 데이터베이스의 적중률 비교

적중률은 고객이 요청한 정보가 검색 유형 데이터베이스에서 찾아질 확률을 말한다. <그림 14>는 트랜잭션별로 검색회수에 따른 적중률 실험을 수행한 결과이다.



<그림 14> 검색회수에 따른 적중률의 변화

트랜잭션 수에 따라 연관 규칙 알고리즘을 적용하여 최소 지지도를 0.5, 신뢰도를 0.7로 하여 검색 유형 데이터베이스를 구축하였다. 이때, 트랜잭션의 수는 누적의 개념으로 사용하였다. 즉, 100개의 트랜잭션은 50개의 트랜잭션을 포함한다.

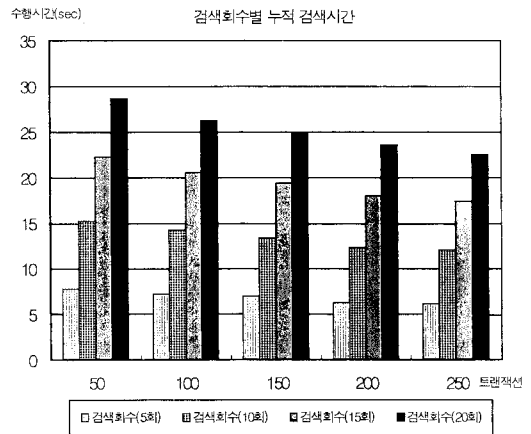
트랜잭션 수(50, 100, 150, 200, 250)에 따라 검색 유형 데이터베이스를 구축한 후 검색회수를 5회부터 5회씩 누적시키면서 20회까지 실험하였다. 결국 20회의 적중률은 15까지의 적중률을 포함한다. 적중률은 트랜잭션별로 검색회수(5, 10, 15, 20)별로 검색 유형 데이터베이스에서 검색 결과를 가져오는 회수를 기록하여 총 검색회수로 나누어 계산하였다.

<그림 14>의 평균 적중률을 살펴보면 트랜잭션의 수와 검색회수가 높아질수록 적중률이 높게 나타나는 것을 알 수 있다. 본 실험에서 평균

적중률은 트랜잭션의 수가 100이상 일 때 0.4를 넘고 있는 것으로 나타났다. 이는 적중률이 높은 편이기 때문에 구축한 상품 검색시스템이 빠른 속도를 가진 시스템이라고 볼 수 있다. 하지만 상대적으로 검색 유형 데이터베이스의 크기는 증가된다.

## 5.3 검색 요청에 따른 검색 시간 비교

인터넷 쇼핑몰 검색 시스템에서 고객이 요청한 검색 정보에 대해 다양하고 정확한 검색 정보를 제공하는 것도 중요하지만, 빠른 검색 정보를 제공하는 것도 매우 중요하다. <그림 15>는 5.2절에서 제시한 검색 유형 데이터베이스 구성과 검색 요청 회수가 동일한 상태에서 원시 상품 데이터베이스 또는 검색 유형 데이터베이스에 접근하여 검색 결과를 응답해 줄 때까지의 누적 시간을 비교한 것이다.



<그림 15> 검색 회수별 누적 검색 시간

<그림 15>에서 20회 입력 검색문을 요청할 경우 200개의 트랜잭션 검색 유형 데이터베이스가 50개보다 전체 접근 시간에서 약 3.79초 감소됨을 보인다. 이때의 적중률은 50개일 경우 0.42이고 200개일 경우 0.58로 증가되었다. 그러므로 트랜잭션 수에 따른 검색 유형 데이터베이스 구성에서 요청한 검색 문항에 대해 적

중률이 높을수록 빠른 접근 시간을 가지기 때문에 누적 접근 시간은 줄어들 수 있었다.

## VI. 결 론

인터넷을 사용하고 있는 대부분의 사용자들은 인터넷 쇼핑물을 통한 전자상거래가 필요하다는데 의심의 여지가 없다. 그러나 인터넷 쇼핑물을 통한 상품 구매 방법은 안전성과 느린 검색 결과 때문에 현재는 많은 사용자를 확보하지 못하고 있다. 하지만 시간 절약, 구매 편리, 가격 저렴 등의 여러 가지 편리함이 제공되고 있기 때문에 향후 많은 사용자들이 참여할 것으로 예상된다. 따라서 본 연구에서는 상품 검색을 위한 편리함과 유용성 그리고 빠른 검색 결과를 제공하는 데 목적을 두었다. 본 연구의 특성과 유용성은 다음과 같다.

첫째, 연관 방법을 적용하여 축약된 검색 유형 데이터베이스를 구축하여 빠른 상품 검색을 지원하는 데이터관리 시스템 구조를 제시하였다.

둘째, 고객이 선호하는 속성별 가중치와 선호 기준을 입력받아 고객이 선호하는 상품을 우선적으로 제공하는 고객 중심의 의사결정 지원 방법을 제시하였다.

따라서, 본 검색 시스템에서는 상품 검색을 빠르게 하고 고객의 선호도에 적합한 상품을 우선적으로 제시함으로써 고객은 상품 선정에 대한 의사결정을 지원 받을 수 있게 된다.

제시한 시스템의 성능을 평가한 실험에서 다

음과 같은 결과를 얻었다.

첫째, 지지도와 신뢰도에 따른 데이터베이스의 크기를 실험하였다. 지지도와 신뢰도가 낮을수록 검색 유형 데이터베이스의 크기는 커졌고 높을수록 데이터베이스 크기는 적어졌다. 이러한 데이터베이스 크기는 적중률에 영향을 미치게 되므로 적절한 지지도와 신뢰도 결정은 중요한 요인이다.

둘째, 트랜잭션의 수와 검색 회수를 높이면서 적중률을 실험하였는데, 적중률이 점점 높아졌다. 이는 적중률이 높을수록 빠른 검색 결과를 제공하지만 상대적으로 검색 유형 데이터베이스의 크기는 증가하게 된다. 또한, 적중률이 높을수록 접근 시간은 줄어들었다.

본 연구의 상품 검색 시스템에는 다음과 같은 연구가 추가되어야 한다.

첫째, 시뮬레이션에서 사용한 상품 검색 데이터보다 더욱 다양한 상품별로 데이터를 추가하여 시뮬레이션을 수행하는 과정이 필요하다.

둘째, 제시한 시스템을 실제 인터넷 쇼핑몰에서 운영하여 수집한 실제 데이터와 실험적인 결과에서 얻은 데이터와 비교 분석해 보는 과정이 필요하다.

셋째, 고객 선호 기반의 다속성 의사결정지원 모델에서 정성적인 속성에 대한 정규화 및 점수 계산 방법을 연구해야 한다.

마지막으로 여러 가지 상품과 고객 분류별로 상품의 연관성을 추출하여 판매 촉진을 도와주는 상품 검색 시스템에 대한 연구가 필요하다.

## 〈참 고 문 헌〉

- [1] 김성희, *의사결정론*, 영지문화사, 1988.
- [2] 이근창, 권오병, 이원준, "지능형 에이전트를 이용한 인터넷 DSS 설계에 관한 연구: 마케팅과 생산관리간의 전략적 통합 문제를 예로 하여," *경영정보시스템 논문지*, 제7권, 제3호, 1997, pp. 1-21.
- [3] 이재규, 최형림, 김현수, 이경전, *전자상거래원론*, 법영사, 1999.
- [4] 황현숙, 어윤양, "전자상거래에서 고객선호 기반의 의사결정 모델 에이전트 시스템에

- 관한 연구," 한국정보시스템학회 논문지, 제 8권, 제2호, 1999, pp. 91-110.
- [5] Agrawal, R., Imielinski, T. and Swami, A., "Mining Association Rules between Sets of Items in Large Databases," *Proceedings of ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.
- [6] Agrawal, R. and Srikant, R., "Fast Algorithms for Mining Association Rules," *Proceedings of the 20th VLDB Conference*, 1994.
- [7] Andriaans, P. and Zantinge, D., *Data Mining*, Addison-Wesley, 1996.
- [8] Art, T., *JDBC Developer's Resource*, Prentice-Hall Inc, 1997.
- [9] Berson, A. and Smith, S., *Data Warehousing, Data Mining and OLAP*, McGraw-Hill, 1997.
- [10] Bhargava, H.K., King, A.S. and Mcquay, D.S., "DecisionNet: Modeling and Decision Support Over the World Wide Web," *Proceedings of the 3rd ISDSS Conference*, 1995.
- [11] Bhargava, H.K., Ramayya K. and Redolf M., "Decision support on demand: Emerging electronic markets for decision technologies," *Decision Support Systems*, Vol. 19, 1998, pp. 193-214.
- [12] Chou S.T., "Migrating to the Web: a Web Financial Information System Server," *Decision Support Systems*, Vol. 23, 1998, pp. 29-40.
- [13] Fayyad, U.M., Gregory P., Padhraic S. and Ramasamy, U., "Advances in Knowledge Discovery and Data Mining," *MIT press*, 1995.
- [14] Gio, W., "Information Systems that Really Support Decision Making," *Journal of Intelligent Information Systems*, Vol. 14, 2000, pp. 85-94.
- [15] Gray, P., "Group Decision Support System," *Decision Support Systems*, Vol. 3, 1987, pp. 233-242.
- [16] Ian H.W. and Eibe F.K., "*Data Mining - Practical Machine Learning Tools and Techniques with Java Implementations*," Morgan Kaufmann Publishers, 1999.
- [17] Guttman, R., Moukas, A. and Maes. P., "Agent-mediated Electronic Commerce: A survey," *Knowledge Engineering Review*, 1998.
- [18] Megiddo, N. and Srikant, R., "Discovering Predictive Association Rules," *Proceedings of the 4th International Conference on KDD and DM*, 1998.
- [19] Morton, S., "Management Decision Systems: Compute Based Support for Decision Making," *Division of Research, Harvard University Cambridge*, 1971.
- [20] Pekowsky, L., *Java Server Pages*, Addison-Wesley, 2000.
- [21] Sarawagi, S., Thomas, S. and Agrawal, R., "Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications," *Data Mining and Knowledge Discovery*, Vol. 4, No. 2, 2000, pp. 89-125.
- [22] Sauter, V., *Decision Support Systems*, Wiley, 1997.
- [23] Sprague, R.H., "A Framework for the Development of Decision Support Systems," *MIS Quarterly*, Vol. 4, No. 4, 1980, pp. 1-26.
- [24] Srikant, R., Imielinski, T and Swami, A., "Mining Associations between sets of Items in Massive Databases," *Proceedings ACM SIGMOD*, 1993.
- [25] Turban, E., *Decision Support Systems and Expert Systems*, Prentice Hall, 1995.

◆ 저자소개 ◆



황현숙 (Hwang, Hyun-Suk)

부경대학교에서 전산학 전공으로 석사, 경영정보 전공으로 박사 학위를 취득하였다. 현재, 부경대학교 강사로 강의하고 있다. 주요 관심분야는 Data Mining, 인터넷 비즈니스, 의사결정지원시스템 등이다.



어윤양 (Eh, Youn-Yang)

서강대학교에서 경영학 석사 및 박사학위를 취득하였다. 현재, 부경대학교 경영학부 교수로 재직중이다. 주요 관심분야는 시스템 모델링과 MCDM 등이다.

◆ 이 논문은 2001년 6월 9일 접수하여 1차 수정을 거쳐 2001년 11월 30일 게재확정되었습니다.