

# 인터넷 비즈니스 기반의 고객관계관리(CRM) 을 위한 웹 로그 분석에 관한 연구

김재형\*, 노효원\*\*, 김남호\*\*, 정정화\*

\*한양대학교 산업대학원

\*\*호남대학교 정보기술원

e-mail : [nhkim@itc.honam.ac.kr](mailto:nhkim@itc.honam.ac.kr)

## A Study on Web-log Analysis for CRM based on Internet Business

Jae-Hyung Kim\*, Hyo-Won Noh\*\*, Nam-Ho Kim\*\*, Jong-Wha Chong\*

\*Graduate School of Industry, Han-Yang University

\*\*Information Technology Center, Ho-Nam University

### 요 약

개별화 웹 마케팅은 본질적으로 고객지향의 패러다임이다. 즉, 개별 고객의 특수한 니즈를 개별적으로 파악해서 각각의 고객에게 차별화된 서비스를 제공하는 것이 그 핵심이다. 웹 서버의 로그파일에 데이터마이닝의 연관규칙 기술을 이용하게 되면 고객행동 패턴의 파악 및 예측을 위한 기법으로 활용할 수 있다. 본 연구에서는 웹 사용자의 교차 판매를 위한 원투원 마케팅에 필요한 접근패턴을 분석하고자 하며, 이는 웹서버 로그파일 분석을 통하여 이루어진다. 분석하고자 하는 웹서버 로그파일은 기존의 데이터웨어하우스의 원천 데이터들과는 다르게 비정형적인 데이터 구조를 가지고 있다. 이들 비정형 데이터 처리와 교차판매 지원을 위한 데이터마이닝 모델링, 이를 통한 원투원 마케팅 모델 제시, 그리고 이의 활용이 고객관계관리(CRM)에 미치는 효과를 제시한다.

### 1. 서론

최근에 기업의 마케팅 환경이 고객관계관리(CRM, Customer Relationship Management)로 대표되는 고객 중심적 마케팅으로의 패러다임 변화가 일고있다.[8] 특히 인터넷은 회사의 상품과 서비스를 마케팅하기 위한 강력한 매개체로 등장하였다. 종전의 대중매체를 통한 마케팅이 광고 주관업체가 불특정 다수의 잠재 고객에게 마케팅하는 One-to-Many의 일방적 형태의 의사소통을 행한 것에 반하여 인터넷 마케팅은 가상의 공간에서 소비자와의 관계형성 및 실시간 상호작용이 가능한 쌍방향 커뮤니케이션을 통한 일대일 개별화 마케팅이 실현될 수 있는 최적의 환경을 제공하고 있다.[5] 개별화 웹 마케팅은 본질적으로 고객지향의 패러다임이다. 즉, 개별 고객의 특수한 니즈를 개별적으로 파악해서 각각의 고객에게 차별화된 서비스를 제공하는 것이 그 핵심이다. 이를 통해서 기업

은 개별 고객과의 관계증진을 통해 고객 유지율을 상승시키고 해당 회사에 대한 충성심을 유지해서 결국은 이익을 증대시킬 수 있는 것이다.

지금까지 개별화 웹 마케팅을 위한 데이터는 사용자들의 회원정보와 통계 데이터로부터 추출되었다. 하지만 웹 서버의 로그파일에 데이터마이닝의 연관규칙 기술을 이용하게 되면 고객행동 패턴의 파악 및 예측을 위한 기법으로 활용할 수 있다.[1]

로그파일 분석의 데이터마이닝 결과는 CRM 주제 중 하나인 교차판매(Cross Selling)에 적용된다. 즉, 한 제품을 구입한 고객이 다른 제품을 추가로 구입할 수 있도록 유도하는 것을 목적으로 한다. 이로부터 기존 고객이 보다 수익성 있는 제품을 구매하도록 하거나 고객의 거래 폭을 넓혀서 이탈을 어렵도록 하여 충성도 향상의 효과로도 연결된다.

본 연구에서는 웹 사용자의 교차 판매를 위한 원

투원 마케팅에 필요한 접근패턴을 분석하고자 데이터 마이닝의 연관규칙 기법을 사용하였으며, 이는 웹서버 로그파일 분석을 통하여 이루어진다. 분석하고자 하는 웹서버 로그파일은 기존의 데이터웨어하우스의 원천 데이터들과는 다르게 비정형적인 데이터 구조를 가지고 있다. 이들 비정형 데이터 처리와 교차판매를 위한 데이터마이닝 모델링, 이를 통한 원투원 마케팅 모델 제시, 그리고 이의 활용이 고객관계관리(CRM)에 미치는 효과를 분석하고자 한다.

논문구성은 1 장에서는 연구배경과 목표에 대하여, 2 장에서는 지금까지 수행되어온 관련연구, 3 장에서는 데이터마이닝을 위한 웹로그 파일 분석에 대해서, 4 장에서는 시스템 설계와 평가에 대해서 그리고 5 장 결론에서는 기대효과 및 향후 연구에 대하여 논하고자 한다.

## 2. 관련연구

### 1) 인터넷 비즈니스 기반의 CRM

CRM 과의 통합을 통해 인터넷 비즈니스는 단지 저렴한 매스 비즈니스 수준에서 정교하게 타겟팅 된 수익성 높은 비즈니스의 모습으로 변화될 수 있다. 이들 간의 연계와 상호 보완적인 기능의 통합은 더 이상 새로운 시장이 생겨나기 어려울 정도로 포화상태에 이른 시장에서 기업의 생존과 지속적인 성장을 위한 통합된 고객중심 마케팅 전략 실행으로의 혁명을 실제로 가능하게 할 것으로 보인다.[7]

효율적인 CRM 을 위해서 꼭 필요한 것 즉, 고객에 대한 특성 정보와 구매 의사결정 기준 그리고 구매 행위에 대한 정보를 획득하고 관리하는 과정에는 상당한 비용과 시간이 소요되며, 그 보다는 최초로 입력되는 고객 정보는 신뢰성이 부족하여, 지속적으로 갱신되지 않으면 CRM 을 통해 얻을 수 있는 이점의 없다. 만일 데이터가 최신의 정보를 정확하게 반영하지 못한다면 이를 기반으로 한 의사결정은 오히려 역효과를 가져올 수도 있기 때문이다.

이러한 데이터의 신뢰성을 제고하고 지속적으로 갱신하기 위한 가장 효과적인 방법이 인터넷의 상호 작용적이고(interactive) 쌍방향적인 커뮤니케이션을 통한 고객 정보 및 반응정보의 관리이다. 이렇게 인터넷 마케팅과 연계되어 구현되는 CRM 은 기존의 CRM 과 근본적으로 다르다. 먼저, 고객 정보를 획득하는 과정과 그 질이 달라지는 측면이다.

인터넷 마케팅은 인터넷에 개설된 홈페이지나 쇼핑몰에 통신망을 통해 고객이 접속하여 인증되면서 시작되는 데, 일단 접속이 되면 고객은 의식하지 못하지만 고객의 모든 행동을 모니터링 할 수 있다.

이렇게 수집된 데이터가 각종 통계, 데이터 마이닝 기법으로 분석되어 고객 개개인을 타겟으로 한 수준 높은 마케팅 조치가 설계될 수 있기 때문에 기존의 CRM 과 그 격이 달라진다.

데이터의 획득과는 반대편에서 인터넷을 통해 얻어낸 고객 개개인의 정보를 활용하는 측면을 살펴보면 앞서 설명한 바와 같이 고객이 인터넷에 회원 확인을

통해 접속하게 되면 데이터베이스에 저장된 개인 마케팅 정보와 그 고객이 전에 구매한 실적, 관심 정보 등이 고객이 요청하는 제품 또는 서비스에 따라 적절하게 배치되어 구매 의사결정을 지원함으로써 궁극적으로는 개인 고객을 목표로 하는 가상 상점을 실시간으로 만들어 낼 수 있다.

물론 이러한 일대일 마케팅을 위한 개인상점이 아직까지는 실현되지 못했지만 첨단 네트워킹 기술과 인공지능 등의 관련 기술들이 끊임없이 개발되고 있으므로 머지 않은 장래에 구현될 수 있을 것이다.

### 2) 웹 데이터 마이닝

어느 웹사이트에 심심해서 방문을 했던지 그렇지 않으면 구체적인 목적을 갖고 방문을 했던지 간에 기업의 웹사이트에 한번이라도 들렀던 사람들이 사이트에서 행한 모든 일은 기록으로 남게 된다. 웹사이트 설정이 어떻게 되었는지에 따라서 다르긴 하지만 방문자가 어떤 경로를 통해 들어왔는지 알 수 있다.

예를 들어, 어떤 검색엔진을 통해 들어왔으며, 어떤 키워드로 검색했는지에 관한 정보들에 대한 로그파일이 남게 된다. 또한 쿠키(사용자 인증에 관한 사용자 브라우저와 서버간의 주고받은 기록)는 홈페이지 방문자의 이동경로 혹은 그 사용자가 이전에 한번 들렀던 적이 있는 사용자인지의 여부를 알려주게 된다. 그러나 그 보다도 더욱 중요한 것은 방문한 고객의 정보가 기록된다는 것이다.[6]

웹사이트에서 수집한 고객등록 정보들과 그 내용들에 대한 마이닝을 통해 기업은 인구통계학적 소비자 선호도를 발견하여 특정 광고나 배너를 포지셔닝 할 수 있도록 하는 기초자료를 추출할 수 있다. 새로운 데이터나 정보가 웹사이트를 통해 수집되면 이 정보들은 지속적으로 데이터웨어하우스로 통합되어 향후의 의사결정에 도움을 주는 분석결과를 제공하고, 데이터베이스 마케팅과 전략기획을 위한 자료로 활용되는 것이다. 또한 웹사이트 데이터 마이닝을 통해서 온라인 상에서 제공하는 서비스와 제품간의 연관관계를 밝혀내어, 적절한 제품이 적절한 서비스와 함께 판매가 되고 있는지의 여부를 밝혀내게 될 수도 있다.

데이터 마이닝을 통해서 기업은 웹사이트상의 패턴을 의미 있는 정보로 종합해내고, 인터넷 상의 고객들과 예상치들을 이해하고 연관시킬 수 있게 된다. 데이터와 웹이 제공하는 방대한 사업지식의 흐름에 근거한 웹 마이닝은 온라인 고객과의 관계를 생성하고 유지시키며 생산성 있는 온라인 상점의 최전선을 구축하는데 있어 결정적 열쇠가 되는 것이다.

## 3. 웹 로그 분석

### 1) 로그분석

로그파일 분석은 사용자가 자신의 사이트에 방문한 경우 로그파일에 흔적을 남기게 되며 이러한 방문자의 정확한 데이터를 기반으로 고객분석을 통하여 마케팅 피드백(Feedback)을 할 수 있는

고객분석 방법이다.[3] 이러한 로그파일 분석을 통하여 주요 고객층, 고객 구매패턴, 주 구매시간, 구매 탐색경로 등의 데이터를 추출할 수 있다. 이러한 데이터를 기반으로 인터페이스 설계나 상품 레이아웃 등의 설계, 고객 서비스의 강화 등의 다양한 대 고객 마케팅을 펼칠 수 있다.

로그파일 분석 결과를 비즈니스에 전략적으로 활용하는 방안은 크게 세가지로 나눌 수 있다.

첫번째는 새로운 신규 사업개발을 들 수 있다. 신규사업 개발은 고객의 다양한 요구를 예측하여 새로운 사이트 개발 및 새로운 시장 기회를 창출하는 것이다.

두번째는 마케팅 및 광고 전략으로서의 활용이다. 기존의 매스마케팅이나 매스광고처럼 무차별적으로 광고집행을 하는 것이 아닌 방문자의 방문 경로 및 페이지뷰(Page View), 임프레션(Impression)이 많은 타겟 페이지(Target Page)에 집중적으로 타겟에 맞는 광고를 집행하여 효과를 보는 것을 말한다.

세번째로 최적의 환경에서 사용자들이 사이트를 탐색하고 방문하도록 서버 및 회선 등의 기술적 자원 및 수행능력 계획을 수립할 수 있다.

로그분석으로부터 얻을 수 있는 분석데이터 정보는 다음과 같이 구분할 수 있다.

▶ 방문자 트래픽(Traffic)정보 - 웹사이트 방문자의 데이터를 토대로 페이지뷰, 임프레션 등의 정보와 방문자들의 트래픽이 가장 많은 시간과 적은 시간의 데이터를 다양한 수치로 비교하여 각 시간대별 방문자를 파악할 수 있으며 방문 형태를 분석하여 사이트 전략 수립에 반영할 수 있다.

▶ 방문경로 정보(Referrers) - 방문자들이 사이트를 찾아오기 위하여 어떠한 경로를 거쳤는지에 관한 데이터로 도메인, 사이트 주소(URL), 뉴스그룹, 이메일, 방문국가, 검색엔진 등과 같은 방문자들의 세세한 방문 경로를 추적하여 사이트 프로모션(Site Promotion)전략수립에 이용 가능하다.

▶ 방문자의 시스템 환경정보(System) - 방문자의 사용 브라우저 환경 및 버전, 모니터 환경, 사용 O/S, 사용언어, 플러그인 등의 접속 환경 데이터를 토대로 방문자들이 최적의 환경에서 사이트 정보를 탐색할 수 있도록 방문자의 환경에 맞게 재구성할 수 있는 데이터를 얻을 수 있다.

▶ 방문자에 관한 방문정보(Visitor) - 방문자가 이용하고 있는 인터넷 서비스 제공업체(ISP), 방문자의 도메인, 방문자 국가정보, 방문 주요 시간대 등의 정보를 토대로 사이트 업데이트 계획을 수립할 수 있다.

▶ 사이트 열람정보(Navigation) - 방문자가 체류한 정보를 토대로 평균 체류시간 및 체류 시간이 가장 많은 사이트 등의 데이터를 토대로 마케팅 및 광고 집행 수행시 자료로서 활용할 수 있다.

▶ 페이지 방문정보(Pages) - 홈페이지 전반에 관한 방문자 정보를 분석하여 가장 많이 방문하는 방문사이트 및 처음 방문하는 페이지(Entry Page), 최종 방문 페이지(Exit Page), 링크 오류 등의 데이터를 분석할 수 있다. 방문 정보는 사이트 오류 수정 및 사이트 인터페이스의 구성 등에 관한 데이터를 제공하여 콘텐츠 재구성시 반영할 수 있다. 이러한 로그분석 데이터 결과를 분석해 나가면 효율적으로 웹사이트 관리 및 비즈니스 전략을 수립할 수 있다.

## 2) 로그 데이터 구조

대부분의 웹 접근 로그는 CERN 과 NCSA 에서 HTTP 프로토콜로 규정한 Common Log Format 를 따른다. 이 표준을 따르는 로그 항목에는 사용자의 IP 주소, 사용자 ID, 접근 시간, 요구 방법, 접근된 페이지의 URL, 데이터 전송을 위한 프로토콜, 에러 코드, 전송된 데이터 길이 등이 포함된다.

일반적으로 특정 웹 페이지를 보기 위해 사용자에 의한 요구의 결과로써 다양한 파일들이 접근된다. 예를 들면, 다양한 이미지, 소리, 비디오 파일, cgi 실행 파일, 이미지 맵 파일의 선택 영역, HTML 파일 등이 있다. 다음 <표 1>는 사용자의 요구의 결과로 접근된 파일들의 유형을 보이고 있다.

<표 1> 사용자 접근 파일 종류

파일 종류	내용
<filename>.html <filename>.gif or <filename>.jpg <filename>.map?<x,y> <program>.cgi?<arguments>	요구된 HTML 페이지 이미지 파일 이미지 맵 파일의 x, y 좌표 서버쪽 실행 파일

로그 데이터 분석의 주요 목적은 특정 서버와 관련된 웹 공간에서 다양한 웹 페이지들에 대한 관심 있는 접근 패턴을 찾아내는 것이다. 서버 로그는 과잉 또는 데이터 마이닝과 관련이 없는 항목을 포함하고 있다. 예를 들면, 대부분의 이미지 파일 항목은 관련이 없거나 필요 이상이다. 몇 개의 이미지 파일을 가지고 있는 URL 이 선택 되었을 때 이미지들은 클라이언트에 전송이 되고 이들 파일은 로그 파일에 독립적인 항목으로 기록된다. 우리의 목적을 위해서는 하이퍼링크가 없는 이미지 파일은 다른 페이지를 방문하는데 사용되지 않기 때문에 무시될 수 있다. 사용자가 하이퍼 링크를 선택하게 되면 목적지 URL 이 로그 파일에 기록이 되기 때문에

하이퍼 링크를 가진 이미지 화일은 과잉 정보다. 이미지 맵 화일이나 다른 멀티미디어 지원 화일들도 비슷한 상황이다.

어떤 표출이 사용자가 특정 URL 를 방문하도록 하는데 호의적인가를 결정하는 것이 아니라면, 웹 마이닝에서는 이와 같이 불필요한 로그 항목들을 제거할 수 있다. 이러한 과정을 데이터 필터링이라 한다. 이 과정에서는 URL 이름의 접미사를 체크 한다. Gif, jpeg, jpg, map 등의 접미사가 붙은 화일명을 가진 URL 항목은 제거된다. 여기서는 실행화일에 대한 요구는 고려하지 않지만 cgi 프로그램은 실행 시점에 HTML 화일을 생성하므로 중복되거나 관련 없는 항목들은 아닐 것이다.

데이터 필터링 과정은 데이터마이닝을 위해 관련된 항목들만을 선택하는 것을 의미한다. 관심 있는 항목들은 IP 주소, 사용자 id, 접근 시간, URL 이다. 로그 항목들이 필터링되면 로그 데이터는 연관 규칙 탐사를 위한 데이터 모델에 알맞은 형태로 변환된다. 데이터 모델은 다음절에 기술되어있다.

### 3) 연관규칙을 위한 데이터모델

하나의 트랜잭션이 자연스럽게 정의되는 바코드 분석과는 다르게 웹 환경하에서는 연관규칙을 찾기 위한 트랜잭션의 정의가 자연스럽게 않다. 사용자에게 의해 주어진 최대 시간 간격 안에 같은 사용자 (IP 주소, 사용자 ID)에 속하는 로그 엔트리를 기초하여 트랜잭션을 정의하였다.  $L$  을 서버 접근 로그 엔트리 집합이라고 한다면, 하나의 로그 엔트리  $l \in L$  은 다음과 같은 항목을 갖는다.

- $l.ip$  : 사용자의 IP 주소
- $l.uid$  : 사용자 ID
- $l.url$  : 사용자가 접근한 페이지의 URL
- $l.time$  : 접근 시간

웹 로그 엔트리에는 요구 방법(POST, GET), 전송 화일 크기 등과 같은 다른 항목들도 있지만, 우리의 목적을 위해서는 위의 필수적인 항목만을 선택하였다.

**정의 1** : 하나의 연관 트랜잭션  $t$  는 다음과 같이 3 개의 항목으로 이루어진다

$$t = \langle ip_b, uid_b, \{l_1.url, \dots, l_m.url\} \rangle$$

이때,  $1 \leq k < j \leq m$ ,  $l_k \in L$ ,  $l_k.ip = ip_t$ ,  $l_k.uid = uid_t$ ,

$$l_j.time - l_k.time \leq \text{최대 시간 간격}$$

**정의 2** :  $URL_t = \{l_1.url, \dots, l_m.url\}$  일때, 연관 트랜잭션 집합  $T = \langle ip_t, uid_t, URL_t \rangle$  가 된다.

웹 공간(web space)은  $WS = \cup_{t \in T} URL_t$  로 정의된다.  $U$  를  $URLs(U \subseteq WS)$ 의 집합이라고 하면,  $U$  에 대한 지지도 (support count)는

$$\sigma(U) = |\{t \mid U \subseteq URL_t\}|$$

즉,  $U$  에 대한 지지도는 한번의 트랜잭션안에서 URL 들이 사용자에게 의해 접근된 횟수이다.

**정의 3** : 연관규칙은  $X \xrightarrow{s, \alpha} Y$  형태로 표현된다. 이때  $X \subseteq WS$ ,  $Y \subseteq WS$ 이다.

$$\text{지지도 } s = \sigma(X \cup Y) / |T|$$

$$\text{신뢰도 } \alpha = \sigma(X \cup Y) / \sigma(X)$$

연관규칙 발견의 문제는 최소 임계치  $s, \alpha$ 가 주어졌을 때,  $X \xrightarrow{s, \alpha} Y$ 의 모든 규칙을 찾아내는 일이다.

### 4) 로그분석을 이용한 타겟 마케팅

로그분석으로부터 나온 결과를 어떻게 인터넷 마케팅 전략에 이용할 수 있는지 살펴본다.

#### ▶ Domain Oriented Targeting

로그화일을 분석하게 되면 어떤 도메인에서 주로 접속하고 있는가를 쉽게 알 수 있다. 그렇게 되면 Focused 된 이용자를 대상으로 타게팅 광고를 실행할 수 있다. 예를 들어, 어느 사이트의 이용자 도메인이 ac.kr, or.kr, co.kr 등이 각각 30%, 50%, 20% 였다고 가정해 보자. 특정 광고주에게는 학생은 전혀 고객으로서 의미가 없는 대상이라고 하면, 광고 집행 중 30%는 의미없는 활동이었다고 평가할 수 있다. 따라서 사전에 사이트의 이용자를 분석해 보고, 해당자만을 취하는 광고활동을 집행할 수 있는데, 이것이 바로 Domain Oriented Targeting 이다. 실제로도 이런 방식의 타게팅은 다양한 광고주에 의해 국내에서도 광범위하게 활용되고 있다.

#### ▶ User Oriented Targeting

단어 의미 그대로 인터넷 이용자 개개인을 중심으로 타게팅이 이루어지는 방식이며, 더 나아가서는 이용자의 성향, 행동을 판단 근거로 타게팅이 가능하다고 하겠다. 이는 회원제 사이트에서 최근의 정보를 이용하여 배너광고나 이메일 푸시를 통한 마케팅 방법이다.

#### ▶ Individual Preference Targeting

이러한 방식의 타게팅은 개인이 제공한 프로파일과 개인의 선호, 취향을 조합하여 타게팅의 정밀도를 높이는 방식이다. 타게팅의 예로는 개인의 연령, 거주지, 성별, 학력수준, 소득, 취미, 여가, 해외여행 경험을 근거로 하는 타게팅이 있겠다.

#### ▶ Behavior Based Targeting

인터넷 마케터 내지 기획자라고 불리는 이들이 궁극적으로 지향하는 타게팅의 왕도는 무엇일까? 바로 실제 구매자들이 보이는 구매 행태를 개개인별로 파악할 수 있고, 그들의 행동반경 내에 늘 우리의 제품을 위치시키고, 유인할 수 있다면

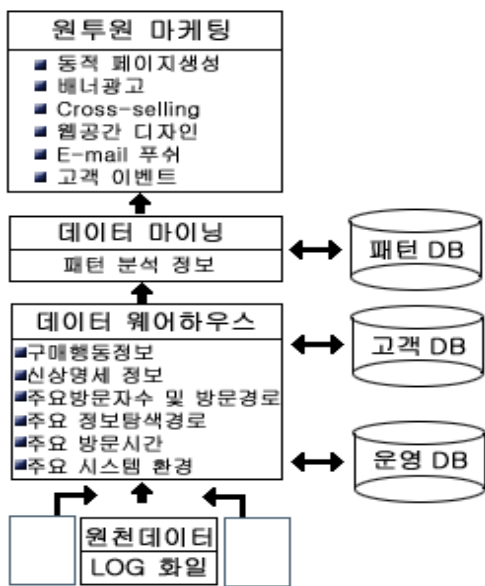


그것이 바로 정답인 것이다. 소비자 구매 행동의 추적 및 이를 활용한 타게팅을 현 수준에서는 아마남의 생각을 읽는 독심술 이상으로 무시무시한 무기로 활용할 수 있을 것이다. 앞으로의 쇼핑몰들이 일반적인 미디어와 차별화 될 수 있는 강력한 파워는 바로 구매를 예측할 수 있다는 점이고 이는 일반 인터넷 미디어들이 보완해야 할 부분이다.

#### 4. 시스템 설계

##### 1) 시스템 구조

본 논문에서 제안한 원투원 마케팅을 지원하는 시스템 구조는 다음 <그림 1>과 같다.



<그림 1> 시스템 구조

##### ▶ LOG 파일 처리

CRM을 위한 분석용 데이터베이스를 구축하는 과정에서 첫번째 작업은 인터넷 사용자의 점점 데이터인 로그파일 항목들 중 분석에 필요한 항목들을 통합하는 모델을 작성하는 것이다. 이를 원시데이터의 전처리라고 하며 사용자 로그파일로부터 관심 있는 항목들인 IP 주소, 사용자 ID, 접근시간, URL를 추출한다. 로그 항목들이 필터링 되면 로그 데이터는 연관규칙 탐색을 위한 데이터모델에 알맞은 형태로 변환된다.

##### ▶ 데이터웨어하우스 구축

고객의 구매 행동 패턴 파악에 가장 활용도가 높은 방법은 기존의 통계적인 프로파일링이다. 여기에서는 고객의 나이, 성별, 거주지, 직업 등 신상명세정보, 사이트에 대한 주요 방문자수 및 방문경로, 주요 방문시간, 주요 시스템 환경을 알 수 있다. 이 방법은 과거의 구매 행동을 단순한 방법으로 일반화 시키는 것이므로 직관적이며 이해가 쉽고 결과 도출을 위해 필요한 노력의 양도 적다. 그러나 해결하고자 하는 개별적인 문제가 가지고 있는 특성에 적합한 지식 또는 해결책을 찾기 위해서는 보다 정교하고 다양한 기

법을 적용하는 것이 필요하다.

즉, 구매패턴 분석을 위한 데이터마이닝에서는 로그 파일로부터 사용자별로 웹 페이지 접근을 개별적인 세션으로 분리한다. 세션을 분류할 때는 타임아웃 방식을 사용하여 제한시간을 초과하였을 때 사용자가 새로운 세션을 시작한다고 볼 수 있다. 이때의 각 세션으로부터 각 사용자에 대해 의미 있는 관계들의 집합을 생성하기 위해 최대 시간간격을 이용하여 세션을 여러 트랜잭션으로 분류하여 탐색경로 데이터베이스에 저장한다.

##### ▶ 데이터마이닝

패턴분석기에는 전처리의 트랜잭션에 포함된 웹 페이지가 항목으로 입력된다. 이 웹 페이지에 연관규칙 (Association Rule)을 적용하여 웹 페이지들 간의 패턴지식을 찾아내고 패턴 데이터베이스에 저장한다. 또 고객이 등록시 저장한 고객 데이터베이스의 고객 정보에서 데이터마이닝의 분류(Classification)기법을 사용하여 지식을 생성하여 저장한다.

##### ▶ 원투원 마케팅 서비스

데이터마이닝 결과의 패턴 DB를 이용하면 교차판매, 동적 페이지 생성, 웹공간 디자인, 이메일 푸쉬 등의 일대일 마케팅을 지원할 수 있다.

대표적인 CRM 주제인 교차판매(Cross Selling)는 한 제품을 구입한 고객이 다른 제품을 추가로 구입할 수 있도록 유도하는 것을 목적으로 한다. 이로부터 기존 고객이 보다 수익성 있는 제품을 구매하도록 하거나 고객의 거래 폭을 넓혀서 이탈을 어렵도록 하여 충성도 향상의 효과로도 연결된다.

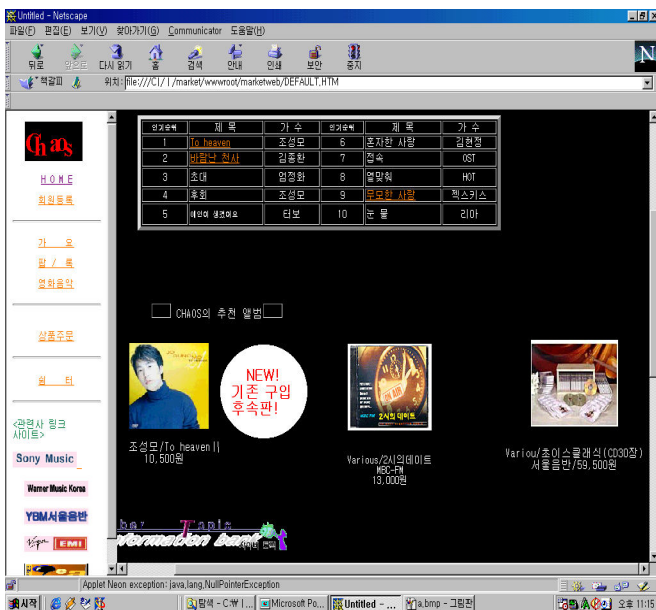
동적 페이지 생성 기능을 이용하게 되면 고객이 쇼핑몰에 접근하였을 때 패턴 DB에 저장된 규칙을 적용하여 고객의 개인별 성향에 맞는 웹 페이지를 동적으로 구성한다. 이는 고객의 방문이 이루어진 이후의 시점에서 동일 고객의 다음 방문에 대한 제안 내용을 결정하는 문제로 웹공간 디자인, 배너광고, 고객이벤트, 교차판매 등과 연동이 되어야 한다.

E-Mail은 고객과 가장 긴밀하게 연결되고 신속하고 빠르게 고객의 욕구를 채워줄 수 있는 직접적인 커뮤니케이션 통로로서 고객 대응의 최접점이다. 패턴분석을 통한 고객에 관한 철저한 분석을 토대로 고객의 욕구에 빠르게 대응하여 고객들이 만족하도록 할 수 있다. 먼저 고객이 입력한 데이터를 토대로 메일로 광고주들이 원하는 타겟 메일(Target Mail)이 발송 가능하도록 업종, 연령, 성별, 연수입, 직업, 지역 등에 따라 세세하게 분류하여 기업이나 광고주들의 타겟 마케팅이 가능하도록 한다. 또한 생성된 패턴 DB를 이용하여 기존 고객과 연관성 있는 정보를 E-Mail로 제공하여 고객과의 지속적인 관계를 유지하여 고객의 충성도를 높이는 마케팅으로 활용할 수 있다.

## 2) 시스템 평가

활용성을 중심으로 제안한 시스템을 평가하고자 한다. 웹사이트를 통한 소비자 구매 행동의 추적 및 이를 활용한 원투원 마케팅은 현 수준에서는 아마 남의 생각을 읽는 독심술 이상으로 막강한 마케팅 무기로 활용할 수 있을 것이다. 인터넷 사이트 운영자와 쇼퍼몰들이 일반적인 미디어와 차별될 수 있는 강력한 힘은 바로 소비자의 행동패턴을 통한 원하는 정보나 구매를 예측할 수 있다는 점이고 이는 지금까지의 인터넷 미디어들이 보완해야 할 부분이다. 구매 행동 활용을 통한 타겟 마케팅이란 어떤 것인지 제안한 시스템의 예를 들어 살펴보자.

기념일을 맞이하여 쇼퍼몰의 꽃배달 서비스를 통해 장미패키지를 구매한 사람은 동일한 연인에게 보낼 샤넬 향수를 화장품코너에서 구매하고, 또 뮤직코너에서 좋아하는 가수의 CD를 구매했다고 한다. 이 사람에게서는 향후 구매했던 아이템을 근거로 일정 형태의 구매 패턴을 갖고있는 것이 확인되므로 그에 맞는 제안형 Proposal을 제시함으로써 직접 매출을 유발할 수 있을 것이다. 아래의 <그림 2>는 사용자가 재방문하였을 때 이전 구매했던 가수의 신집 앨범을 동적으로 자동 소개해주는 화면이다. 이와 같은 것이 바로 구매 행동을 기반으로 한 원투원 마케팅인 것이다.



<그림 2> 동적인 원투원 서비스

## 5. 결론

연관규칙을 이용한 데이터마이닝 기술을 웹 로그 데이터에 적용하여 인터넷 사용자의 구매행동을 파악하였다. 로그 데이터 처리를 위한 모델링과 시스템을 제안하였으며, 생성된 규칙을 이용한 다양한 마케팅 기법들을 소개하였다. 지금까지의 로그분석 툴들이 제시한 사이트의 방문자가 누구이며,

가장 많이 또는 적게 접근한 페이지 분석 그리고 방문자가 어디로부터 오는지, 현재 사이트의 접근량은 얼마되는지 등의 한계점을 극복하였다.

새로운 방문자에 대한 예측 분류나 행동 양식에 대한 예측에 기초하여, 우리는 비로소 그 사람에게 가장 적절한 광고와 마케팅 메시지들을 전달할 수 있게 된다. 즉, 그 사람이 어떤 상품을 구매하고 싶어하는지를 먼저 예측하여 그 사람에게 적절한 상품 판촉 메일을 보낼 수 있게 되는 것이다. 이는 인터넷이 기업의 대고객 접점으로서 중요성이 날로 중요시 되는 현시점에서 쇼퍼몰이나 웹사이트 운영 기업의 고객관계관리(CRM)를 위한 원투원 마케팅의 중요한 기초 데이터로 활용될 수 있다.

본 논문에서 제안한 시스템은 쇼퍼몰을 운영하는 기업의 입장에서 지속적인 고객관리를 지원하는 시스템이다. 하지만 인터넷을 중심으로 한 디지털 경제에서는 소비자가 가격을 결정하는 등 경제의 주체가 소비자에게로 넘어가고 있다. 이에 따라 향후에는 고객을 직접 지원해주는 에이전트에 관한 연구가 필요하다.

## 참고문헌

- [1] Dong-Ha Lee, Dong-Yal Seo, Nam-Ho Kim, and Jeon-Young Lee, "Discovery and Application of User Access Patterns in the World Wide Web," 4<sup>th</sup> World Congress on Expert system 98, March 16-20, Mexico City
- [2] Bamshad Mobasher, Cooley, J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", in Journal of Knowledge and Information Systems, Vol. 1, No. 1, 1999.
- [3] Bamshad Mobasher, N. Jain, E. Han and J. Srivastava, "Web Mining: Pattern Discovery from World Wide Web Transactions", Technical Report TR96-050, Department of Computer Science, University of Minnesota, 1996
- [4] R. Srikant, R. Agrawal. "Mining Generalized Association Rules", Proc. of the 21st Int'l Conference on Very Large Databases, Zurich, Switzerland, Sep. 1995.
- [5] 안창모, "맞춤 마케팅과 인터넷광고", [http://www.diamond.co.kr/sabo/sabo\\_main\\_sub\\_bot1\\_2.htm](http://www.diamond.co.kr/sabo/sabo_main_sub_bot1_2.htm).
- [6] WEBLOG 사, <http://www.weblog.com/kr/faq.html>
- [7] 방성희, "CRM이란 무엇인가", 월간 경영과컴퓨터.
- [8] 한재홍, "CRM 특강", <http://www.dpc.or.kr/dbworld/document/9912/tech-1.html>
- [9] e.g. Software Inc. Webtrends. <http://www.webtrends.com>, 1999.