

데이터마이닝 기법을 적용한 지능형 인터넷 쇼핑몰 설계 및 구현

김진철^o, 황보윤^{*}, 황성희^{*}, 곽난희^{*}, 문현정^{*}, 우용태^{*}

* 창원대학교 전자계산학과 데이터베이스연구실

A Design and Implementation of Intelligent Internet Shopping Mall using Datamining Technology

Jin-cheol Kim^o, Youn Hwangbo^{*}, Seong-hee Hwang^{*},

Nan-hee Kwak^{*}, Hyeon-jeong Mun^{*}, Yong-tae Woo^{*}

* Dept. of Computer Science, Changwon National University

요 약

본 논문에서는 오라클사의 전자상거래용 패키지인 ICS를 기반으로 지능형 인터넷 쇼핑몰을 설계하고 구현하였다. 본 쇼핑몰은 고객들의 접속 기록과 상품 구매 기록을 데이터마이닝 기법에 의해 통계적으로 분석하여 상품에 대한 인기도에 따라 상품 진열을 자동적으로 구성할 수 있는 지능형 쇼핑몰이다. 본 시스템을 통하여 쇼핑몰 관리자의 주관적인 판단에 의해 수 작업으로 이루어지는 기존 쇼핑몰 관리 업무를 자동화할 수 있다. 또한 최근에 급격하게 증가하고 있는 전자상거래 시장에서 경쟁력을 강화할 수 있는 새로운 형태의 마케팅 기법을 제시하였다.

1. 서론

인터넷을 통하여 시간과 공간의 제약없이 상품을 소비자에게 직접 판매하는 전자상거래 시장이 급속도로 확산되고 있다. 1999년 현재 전세계에서 전자상거래 시장 규모는 39억 9천만달러 정도이며 2003년에는 1조달러 정도로 예상되고 있다[1]. 전자상거래의 성공 사례중의 하나로 2,100명의 직원을 가진 세계 최대의 인터넷 쇼핑몰인 아마존서점의 연간 매출액이 7억 달러이다. 이는 3만명의 직원을 가진 미국 최대 서점인 반스&노블의 매출액인 연간 6천만달러를 훨씬 능가하고 있다[2]. 국내에서도 약 820개의 쇼핑몰이 구축되어 활발하게 전자상거래가 이루어지고 있다[3].

한편 기업에서 효과적이고 신속한 비즈니스 전략과의사 결정을 위하여 데이터마이닝 기법이 널리 활용되고 있다. 데이터마이닝 기법은 기존에 알려진 정보

뿐만 아니라 전혀 예상하지 못하고 쉽게 드러나지 않는 정보까지 대용량 데이터베이스로부터 찾아내고자 하는 정보추출 방법론의 하나이다[4]. 특히 고객들의 상품 구매 취향이 다양해짐에 따라 데이터마이닝 기법을 이용하여 개인별 특성을 고려한 일대일마케팅 전략을 수립하는데 이용되고 있다.

본 논문에서는 오라클사의 ICS(Internet Commerce Server)를 기반으로 데이터마이닝 기법을 적용한 경남닷컴 인터넷 쇼핑몰을 설계하고 구현하였다. 제안한 시스템은 인터넷 쇼핑몰의 마케팅 효율을 개선하기 위한 새로운 전략으로 고객들의 쇼핑몰 접속 기록과 구매 기록을 데이터마이닝 기법에 의해 분석하여 동적으로 화면을 구성할 수 있는 지능형 쇼핑몰이다. 이 시스템에서는 사용자들의 쇼핑몰 접속 빈도와 구매 빈도를 일정 시간동안 축적한 후, 데이터마이닝 기법에 의해 상품별로 가중치를 부과하여 쇼핑몰 화면을 동적으로 구성할 수 있다. 제안된 시스템을 통한

여 소비자들에게 인기가 있는 상품이나 판매 실적이 우수한 상품을 우선적으로 출력할 수 있다. 경남닷컴은 경상남도에서 우수 상품으로 추천한 QC(Quality Certificate) 상품을 인터넷을 통하여 판매하기 위한 인터넷 쇼핑몰이다.

2. 경남닷컴 인터넷 쇼핑몰

2.1 경남닷컴의 개요

미래의 상거래는 인터넷을 통한 전자상거래를 주축으로 전개될 것으로 예측되고 있으며, 많은 기업들이 인터넷 쇼핑몰을 경쟁적으로 구축하고 있다. 특히 인터넷 쇼핑몰은 고객들이 시간과 거리에 제한을 받지 않는 특징으로 인하여 쇼핑몰을 운영하는 기업들은 경쟁력 강화를 위한 새로운 전략 개발에 심혈을 기울이고 있다. 이미 국내의 쇼핑몰에서는 경품 제공이나 부가서비스 제공을 통한 고객 확보 경쟁에 나서고 있으며, 기존의 포털 사이트에서는 기 확보된 사용자들 대상으로 전자상거래를 시도하고 있다.

본 논문에서는 아마존서점에서 채택한 오라클사의 ICS를 이용하여 경남닷컴 인터넷 쇼핑몰을 설계하고 구현하였다. 경남닷컴은 경상남도에서 지역 상품의 판로개척을 위해 우수 상품으로 선정한 QC 상품을 판매하기 위한 인터넷 쇼핑몰이다.

다음 그림 1은 본 논문에서 구현한 데이터마이닝 기법을 적용한 경남닷컴 인터넷 쇼핑몰의 기본 구성도이다.

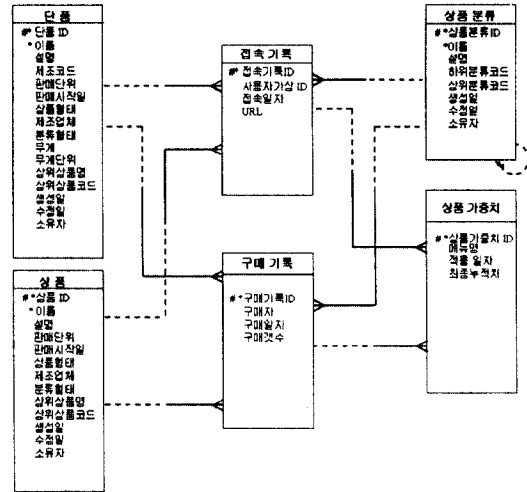


그림 2. ICS를 기반으로 한 경남닷컴 인터넷 쇼핑몰 데이터베이스 설계

단품, 상품, 상품분류 테이블은 ICS에서 기본적으로 제공되는 테이블이다. 그리고 접속 기록, 구매 기록, 상품 가중치데이터는 데이터마이닝 기법을 적용하기 위해 별도로 설계한 테이블이다.

2.3 사용자 접속 기록 분석

최근에 웹 서버에 기록된 로그 파일을 이용하여 사용자들의 홈페이지 접속 패턴을 분석하거나 웹 서버의 부하를 분석하기 위한 연구가 진행되었다[5, 6]. 또한 접근 빈도에 따라 실시간으로 뉴스 데이터를 동적으로 배치하기 위한 연구도 제시되었다[7]. 일반적으로 웹 서버에 대한 접속 기록은 CERN과 NCSA에서 HTTP 프로토콜의 일부로 명시된 Common Log Format을 따라 로그 파일을 생성한다. Common Log Format으로 된 기존 웹 서버의 로그 파일은 접속하는 홈페이지내의 내용과 이미지에 대한 접속자 IP주소, 접속자, 접속시간, 요청문서, 접속프로토콜, 접속상태, 전송된 데이터 크기 등과 같은 정보를 기록한다. 다음 그림 3은 Common Log Format에 의한 로그 기록의 예이다[8].

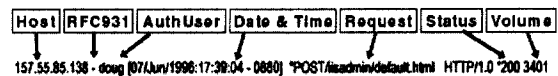


그림 3. Common Log Format 및 로그 기록의 예

그러나 Common Log Format에 의한 접속 기록은 로그 파일의 용량이 지나치게 커지고 실시간 처리가 어려운 문제점이 있다. 본 시스템에서는 고객들의 접속 기록을 실시간으로 분석하기 위하여 쇼핑몰에 접속하는 URL 정보를 전처리를 통하여 데이터베이스

그림 1. 경남닷컴 인터넷 쇼핑몰 시스템 구성도

2.2 경남닷컴의 데이터베이스 설계

본 논문에서 구현한 경남닷컴 쇼핑몰시스템은 오라클사에서 개발한 전자상거래용 패키지인 ICS를 이용하여 구현하였다. 그림 2는 경남닷컴 쇼핑몰 구축을 위한 데이터베이스 설계도이다.

에 저장하는 방식을 채택하였다. 그리고 구매 기록은 별도의 구매 기록 테이블에 저장된다.

2.4 접속 기록 분석을 위한 데이터마이닝 기법

그 동안 데이터마이닝 기법을 이용하여 인터넷 쇼핑몰 사용자의 전공, 직업, 종교, 취미 등과 같은 신상 정보나 구매 패턴에 따라 개인별로 동적으로 화면을 구성하기 위한 연구가 진행되었다[9]. 그러나 대부분의 인터넷 쇼핑몰에서는 로그 인에 따른 사용자의 불편을 해소하기 위하여 상품을 구매하는 시점에서 로그 인을 하는 방식을 채택하고 있다. 따라서 사용자가 상품을 구매하기 전에는 개인별 특성에 따라 동적인 화면을 구성하기 어렵다.

본 논문에서는 데이터마이닝 기법에 의해 인터넷 쇼핑몰에 전시된 상품에 대한 사용자들의 접속 기록을 분석하여 일정 시점마다 동적으로 화면을 구성할 수 있는 지능형 쇼핑몰 시스템을 개발하였다. 이 시스템을 통하여 사용자들이 가장 관심을 가지는 메뉴나 인기 상품을 우선적으로 출력하여 인터넷 쇼핑몰의 마케팅 효율을 개선할 수 있다.

다음 그림 4는 본 시스템에서 적용한 데이터마이닝 기법에 대한 개념도이다.

그림 4. 경남닷컴에서의 데이터마이닝 시스템 개념도

그림 4에서 전처리기는 사용자 접속 기록으로부터 쇼핑과 관련된 정보만을 추출하여 접속 기록테이블과 구매 기록테이블에 저장하는 기능을 수행한다. 접속 기록테이블에 저장되는 정보는 고객들이 인터넷 쇼핑몰을 접속할 때마다 발생하는 접속기록 ID, 상품 코드, 단품 코드, 상품 분류코드, 사용자 가상 ID, 접속일자 및 URL이다. 여기서 사용자 가상 ID는 고객이 구매를 위한 신상정보를 입력하기 전에 임시 저장 공간인 장바구니에 저장하고 다른 상품을 추가 구매 또는 구매 취소시에 해당 고객을 구별하기 위하여 시스템에서 순차적으로 부여한 고유 ID이다. 구매 기록

테이블에는 구매 기록 ID, 상품 코드, 단품 코드, 상품 분류 코드, 구매일자, 구매 개수, 구매자 등과 같은 구매와 관련된 정보가 저장된다.

데이터마이닝시스템은 인터넷 쇼핑몰에서 고객들의 접속 패턴이나 구매 패턴을 분석하기 위한 시스템이다. 먼저, 전처리기에서 생성된 구매 기록테이블과 접속 기록테이블로부터 상품별로 가중치를 부여하여 통계분석 후 가중치 테이블에 저장된다. 구매 기록과 접속 기록간의 가중치는 백화점 매출액 분석시 사용하는 파레토의 80:20 법칙에 근거하여 생성하였다[10]. 여기서 파레토 법칙이란 이탈리아의 경제학자였던 파레토가 주장한 법칙으로 상위 20% 사람이 80%의 부를 소유하게 된다는 경험적 법칙을 의미한다. 이 법칙을 인터넷 쇼핑몰에 적용하여 상품을 구매한 고객은 단순히 상품을 구경한 고객보다 관심도가 4배 높은 것으로 가정하였다. 따라서 구매 기록과 접속 기록간에 4:1비율로 가중치를 부여하였다. 각 상품별 가중치는 <상품접속횟수> + <상품구매수량>×4 로 구하여 가중치 테이블에 저장된다.

마지막 단계로 일정 기간동안 누적된 상품별 가중치를 내림차순으로 정렬하여 쇼핑몰 출력 화면의 우선 순위를 결정하게 된다. 이러한 방법은 고객들이 가장 관심을 가지는 인기 상품을 우선적으로 출력하여 동적으로 쇼핑몰을 구성할 수 있다.

3. 실험 환경 및 구현 방법

본 경남닷컴발의 구현 환경은 Ultra 1, 오라클 DBMS 7.3, 오라클 Internet Commerce Server 1.1, 오라클 Application Server 3.0을 이용하였다. 그리고 응용 프로그램은 PL/SQL과 JAVA 언어로 개발하였다.

3.1 경남닷컴발의 초기 화면 구성

경남닷컴발의 초기 화면은 메뉴 구성을 단순화하여 사용자들이 2~3번의 접속으로 상품 구매 단계까지 도달할 수 있도록 사용자 편의성을 고려하였다. 그리고 키워드에 의해 상품을 검색할 수 있는 기능도 함께 제공하였다.

경남닷컴발에서는 경상남도에서 우수 상품으로 추천한 QC 상품을 인터넷을 통해 판매하기 위한 전자상거래 시스템의 Prototype 모델이다. 현재 경상남도에서는 지역 상품의 판로개척을 위해 116개 업체에서 생산한 농산물, 수산물, 축산물, 공예품, 공산품 분야의 162개 품목을 QC 상품으로 지정하여 통신 판매를 하고 있다.

초기 화면은 162개 QC 상품을 주제별로 7개의 메뉴로 분류하여 구성하였다. 주제별 메뉴는 만물가게, 선물가게, 알뜰가게, 테마가게, 추천상품, 할인상품, 주문예약으로 구성되어 있다. 그림 5는 본 연구에서 구현한 경남닷컴발의 초기 화면이다.

4. 결론

본 논문에서는 오라클사의 전자상거래용 패키지인 ICS를 이용하여 지능형 인터넷 쇼핑몰을 설계하고 구현하였다. 본 시스템에서는 고객들의 접속 기록과 상품 구매 기록을 데이터마이닝 기법에 의해 통계적으로 분석하여 상품에 대한 인기도에 따라 상품 진열을 동적으로 구성할 수 있는 방법을 제시하였다. 고객들의 접속 기록과 상품 구매 기록에 대한 통계적인 분석을 위해 별도의 테이블을 설계하여 기존 ICS에서 제공하는 테이블과 결합하였다. 제시된 방법에 의해 쇼핑몰 관리자의 주관적인 판단에 의해 수 작업으로 이루어지는 쇼핑몰 관리 업무의 일부를 자동화할 수 있다. 또한 최근에 급격하게 증가하고 있는 인터넷 쇼핑몰 시장에서 경쟁력을 강화할 수 있는 마케팅 기법을 제시하였다.

앞으로 고객들의 접속 기록과 구매 기록으로부터 상품들간의 연관성을 분석하여 관련 상품들끼리 그룹화하여 제시함으로써 보다 지능적인 쇼핑몰의 개발과 쇼핑몰 관리를 자동화할 수 있는 방법에 대한 연구를 진행할 계획이다.

참고문헌

- [1] <http://www.nca.or.kr/cgi-bin/CrazyWWWBoard.cgi?db=ncabbs3&mode=download&num=8>
- [2] <http://www.worldsh.co.kr/ec.htm>
- [3] 한국경제신문, "인터넷 쇼핑몰 6.4%만 '흑자' .. 500개사 조사," 1999.08.27
- [4] 한국썬소프트웨어 DB마케팅팀, "데이터마이닝은 향후 정보산업을 이끌어갈 주제," 데이터베이스월드, pp.77~84, 1998.2
- [5] 남도원, 이동하, 서동렬, 이진영, "웹 로그에서의 사용자 접근 패턴 분석", 정보과학회 인간과 컴퓨터상호작용 연구회 회보, 제7권 1호, pp.160~165, 1998
- [6] 김강희, 박노삼, 조은희, 안광선, "웹 서버 접속 로그에서 웹 서버 부하 패턴 분석," 한국정보과학회 봄 학술발표논문집, 제25권 1호, pp.134~136, 1998
- [7] 정귀옥, 박성호, 박용운, 김영주, 정기동, "데이터 접근 빈도를 고려한 실시간 뉴스 데이터의 배치 방법," 한국정보과학회 봄 학술발표논문집, 제25권 1호, pp.36~38, 1998
- [8] Joel Piphagen Alaina Kanfer, Ph.D. "In Search of the Elusive User : Gathering Information on Web Server Access," <http://www.ncsa.uiuc.edu/trg/webstats/>
- [9] 객난희, "WWW기반의 데이터마이닝 모델 설계," 창원대학교 대학원 전자계산학과 석사 학위 논문, 1998.12
- [10] 서울경제신문, 80/20법칙, http://www.korealink.co.kr/13_6/9805/c3651408.htm

그림 5. 경남닷컴의 초기 화면

3.2 데이터마이닝 기법을 적용한 동적 화면 구성

현재 대부분의 인터넷 쇼핑몰에서는 고객들의 관심을 끌기 위한 마케팅 전략의 하나로 주기적으로 첫 화면의 구성을 변경하여 인기 상품이나 이벤트를 소개하고 있다. 그러나 이러한 작업이 쇼핑몰 관리자의 주관적인 판단에 의한 수 작업으로 이루어지는 관계로 쇼핑몰 관리에 따른 추가 비용과 과학적인 근거에 의한 화면 구성이 어렵다.

본 논문에서 구현한 경남닷컴 쇼핑몰에서는 이러한 문제점을 개선하기 위하여 데이터마이닝 기법에 의해 사용자들이 일정 기간 방문하면서 남긴 접속 기록과 구매 기록을 분석하여 초기 화면의 상품 진열을 동적으로 구성할 수 있다. 즉, 각 상품에 대한 접근 기록과 구매 기록의 빈도 수와 가중치를 통계적으로 분석하여 초기 화면에서 상품의 출력 우선 순위를 자동적으로 구성하게 된다. 따라서 고객들에게 과학적인 정보에 근거한 인기 상품을 소개하여 인터넷 쇼핑몰의 마케팅 효과를 기대할 수 있다. 다음 그림 6은 본 논문에서 제안한 데이터마이닝 기법에 의해 초기 화면의 상품 진열이 동적으로 구성되어 출력된 화면의 예이다.

그림 6. 데이터마이닝 기법을 적용한 동적 화면구성