

웹마이닝을 이용한 상품 추천시스템 설계 및 구현

이정호*, 박두순**

순천향대학교 전산학과*, 순천향대학교 정보기술공학부**

A Design and Implementation of Goods Recommendation System using Web Mining

Kyong-Ho Lee*, Doo-Soon Park**

Dept. of Computer Science, SoonChungHyang University*

Division. of Information Technology Engineering, Soonchunhyang University**

요 약

웹의 급속한 성장으로 수많은 양의 정보가 매일같이 쏟아져 나오고 있다. 이는 특정 상품정보를 얻으려는 고객들에게 많은 혼란을 야기할 수 있다. 이러한 문제의 해결을 위해 추천시스템이 개발되었고, 추천 시스템은 고객들이 E-Commerce 상에서 상품을 구매하는 것을 도와주기 위해서 지속적인 증가추세로 사용되고 있다. 이러한 추천시스템은 다양한 고객들의 선호도에 따라 유사성과 비유사성에 대한 정보의 기초위에서 고객들의 잠재적인 관심 항목들에 대해 개인의 취향에 맞게 추천하는 기술들을 제공한다. 그러나, 추천시스템에 많은 관심을 가지면서도 불구하고 그들의 성능에 대한 공개된 기술이나 정보는 매우 제한적이다. 본 논문에서는, 과거 고객들의 구매행동, 고객정보, 데이터마이닝의 연관규칙을 이용한 E-Commerce 추천시스템을 설계하고 구현하였다.

1. 서 론

IT 산업의 급진적인 발전과 인터넷이 폭넓게 보급되면서 웹상에서 얻을 수 있는 정보의 양이 급증하게 된다. 또한, 이러한 웹의 발전은 오프라인에서 이루어지던 상거래를 웹에서 가능하게 하였으며, 이를 전자상거래(Electronic Commerce)라 한다. 전자상거래 기술의 발전으로 수많은 인터넷 쇼핑몰이 웹에 등장하게 되었으며, 인터넷 쇼핑몰들은 저마다 수많은 상품정보들을 자신의 쇼핑몰에 축적하고 있다. 이런 수많은 상품정보로 인해 상품을 구매하고자하는 소비자들은 원하는 상품을 찾기가 매우 어려운 실정에 이르게 되고, 원하는 상품을 찾는데 많은 시간낭비와 노력을 기울여야 하는 것이

현 실정이다.

또한, 대부분의 인터넷 쇼핑몰들은 단순히 상품 판매에만 의존하는 실태이다. 이러한 인터넷 쇼핑몰들은 고객의 수요증가 및 다양성으로 인한 질적인 향상 및 차별화된 서비스를 필요로 하게 된다. 따라서 고객들의 구매패턴 분석, 로그기록 분석, 및 연관규칙을 이용한 데이터마이닝 기법을 통한 상품추천엔진 개발은 단순한 상품판매가 아닌 개개의 고객에게 상품을 자동으로 추천해주는 차별화된 서비스를 가능하게 한다. 이 방법은 고객들의 재구매율 증진 및 고객의 효율적인 관리 뿐만 아니라 향후 인터넷 쇼핑몰의 구조적인 문제의 해결방안에 있어 하나의 해결책이 될 것으로 본다.

본 연구는 한국과학재단 목적기초연구(R05-2002-000-00218-0, 2003) 지원으로 수행되었음.

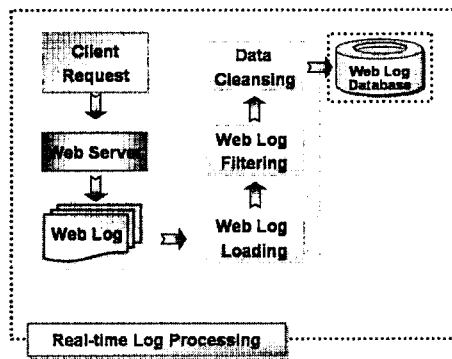
본 논문에서는 상품추천엔진 개발을 위해 다량의 고객정보를 수집해야 하며, 이는 웹로그 수집 모듈개발과 고객의 행위정보 수집 모듈개발로써 고객의 행동 패턴정보를 수집하는데 사용되었다. 또한 고객의 등록정보 및 캠페인 수행정보를 통한 고객 정보수집도 이루어졌다. 이렇게 수집된 다량의 정보들은 상품추천엔진의 전처리과정에서 대부분 수행되게 된다. 상품추천엔진에서 유용하게 사용될 데이터들은 전처리과정을 통해 데이터베이스에 저장되고, 이러한 데이터들과 고객의 구매데이터 및 등록데이터등을 근거로 상품추천 DB가 구축되는 것이다.

이러한 상품추천엔진 개발과 상품추천서비스에 대한 연구[1,2,3,4,5]는 매우 활발하게 진행되었으며, 본 논문에서는 상품추천엔진과 이를 이용한 상품 추천물을 설계하고 구현하였다.

2 상품추천시스템 설계 및 구현

본장에서는 상품추천시스템에 있어 필요한 웹로그처리시스템, 전처리시스템, 패턴분석방법에 대해서 기술한다.

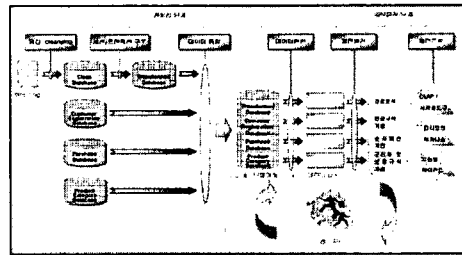
[그림 1]은 웹서버에 수집된 Web Log의 내용을 추출하는 방법을 도식화한 것으로 고객의 행위정보를 추출하는 시스템을 나타낸다[6].



[그림 1] Web Log 처리 시스템

[그림 2]는 본 논문에서 구현한 상품추천시스템

의 전체적인 시스템 구성도로서 전처리과정과 패턴분석과정을 보여주고 있다. 먼저 전처리과정은 데이터 정제(data cleaning), 사용자 구분(session identification), 세션보정(path completion), 트랜잭션 구분(transaction identification) 등으로 구성되어 있다. 그리고, 패턴분석과정에서는 전처리 과정에서 얻어진 정보를 토대로 비슷한 고객들은 묶어주고, 구매한 상품들 간의 연관성을 파악하여 상품 추천 DB를 구성하게 된다.



[그림 2] 전처리 단계와 패턴분석 단계

상품들 간의 연관성 측정의 기본은 얼마나 자주 구매되었는가하는 빈도다. 그리고, 이 빈도를 기반으로 연관정도를 정량화하기 위해서는 다음 세 가지 기준을 고려하게 된다.

- 지지도(Support) - 전체 거래중 항목 X와 항목 Y를 동시에 포함하는 거래가 어느 정도 인가를 나타내주며 전체적 구매도에 대한 경향을 파악할 수 있다.

$$S = P(X \cap Y) = \frac{\text{품목 X와 품목 Y를 포함하는 거래수}}{\text{전체 거래수 } (N)}$$

- 신뢰도(Confidence) - 항목 X를 포함하는 거래중에서 항목 Y가 포함될 확률은 어느 정도인가를 나타내주며 연관성의 정도를 파악할 수 있다.

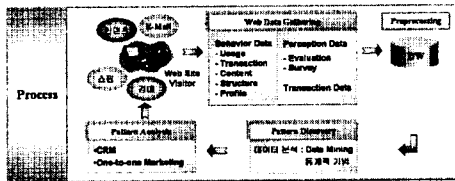
$$C = P(Y|X) = \frac{P(X \cap Y)}{P(X)} = \frac{\text{품목 X와 품목 Y를 포함하는 거래수}}{\text{품목 X를 포함한 거래수}}$$

- 리프트(Lift/Improvement) - 항목 X를 구매한 경우 그 거래가 항목 Y를 포함하는 경우와 항목 Y가 임의로 구매되는 경우의 비를 나타내 준다.

$$L = \frac{P(Y|X)}{P(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$$

본 논문에서는 위에서 말한 지지도, 신뢰도, 리프트를 이용하여 상품들 간의 연관성을 분석하였고, 이를 통해 상품추천시스템을 개발하였다[7].

[그림 3]은 상품추천 엔진을 통한 상품추천이 이루어지는 과정을 도식화한 그림이다. 일반적으로 클라이언트에 해당하는 E-Customer가 상품 추천물에 들어와서 쇼핑을 하고 상품을 구매하는 등의 일련의 과정을 거치게 된다. 또한 상품 추천물은 각각의 고객에 대한 구분을 위해 세션을 활성화하고 고객행위 등을 기록하게 된다. 즉, 세션이 활성화되면 상품추천엔진의 고객의 행위패턴 수집기가 고객의 행위등을 기록하게 된다. 또한 세션이 비활성화되면 수집은 끝나게 된다. 상품추천엔진의 주된 기능으로 이렇게 수집된 기록 등을 토대로 위에서 말한 전처리과정을 통해 상품추천 DB를 구축하게 되는 것이다. 이와같이 상품추천 DB는 실시간 추천시스템에 적용되게 된다. 본 논문에서는 [그림 3]의 전체적인 상품추천흐름도와 전처리과정, 개인화된 정보 추천방법등을 이용하여 상품추천시스템을 구현하였다[8].



[그림 3] 상품추천 흐름도

3 구현 예

본 논문에서는 JAVA 기반 기술을 이용하여 전체 시스템을 구현하였다. 특히, 웹 모듈에 해당하는 부분 즉, 상품 추천물은 JSP 기술을 사용하였

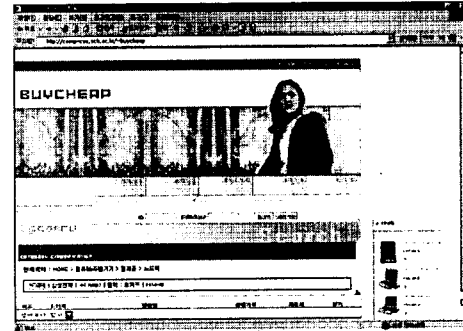
고, 실시간 상품추천엔진 부분은 EJB 기술을 적용하였다. 또한 비즈니스 로직에 관련된 부분은 EJB 컨테이너에서 처리하도록 구성하여 웹서버 성능향상을 고려하여 구현 하였다.

[그림 5]는 상품추천물의 메인화면을 보여주고 있다. [그림 5]의 왼쪽부분에 보여지고 있는 상품카테고리는 부모카테고리와 자식카테고리 형식으로 전체카테고리가 트리구조 형태로 구성하여 표현하였다.



[그림 5] 상품 추천물의 메인화면

[그림 6]은 상품 추천물의 상품진열을 나타내고 있는 화면으로 product_list.jsp에 의해 상품진열이 이루어진다. product_list.jsp는 현재 카테고리에 상품아이템이 없을 경우 자식 카테고리에서 상품추천 리스트를 얻어, 이를 우측에 보여지게 된다.

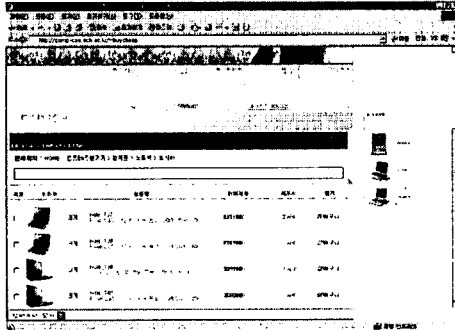


[그림 6] 상품 추천물의 상품진열 1

[그림 7]은 [그림 6]과 같이 product_list.jsp에

의해 보여진다. 하지만 [그림 6]과는 다르게 여러 개의 상품들이 진열되어 있는 것을 확인할 수 있다. 이와같이 현재 카테고리에 상품들이 존재하게 되면 현재 카테고리에 있는 상품들 중에서 고객에게 알맞은 상품을 추천해 준다.

또한 고객이 상품을 구매하기 위해 장바구니에 상품을 담게 되면 장바구니에 담겨져 있는 상품리스트들을 대상으로 상품추천이 이루어진다. 즉, 고객이 장바구니에 상품을 담지 않고 있을 경우에는 현재 카테고리에 대한 관심도만으로 상품을 추천하고, 상품을 장바구니에 담게 되면 장바구니에 담겨져있는 상품을 대상으로 상품추천이 이루어진다.



[그림 8] 상품 추천물의 상품진열 2

4 결 론

본 논문에서는 상품추천엔진을 설계 및 구현하기 위해 데이터마이닝의 연관규칙기술을 이용한 상품추천 알고리즘을 개발하였고, 이를 웹로그 추출 모듈설계 및 실시간 E-Customer 행위 추출 모듈을 개발하였다. 이는 상품추천엔진을 구성하는데 필수적인 요소들로 이러한 요소들을 통해 상품추천엔진개발과 이를 상품추천몰에 적용하여 상품추천서비스를 구현하였다.

참 고 문 헌

[1] J. B. Schafer, J. A. Konstan, and J. Riedl, "E-Commerce recommendation

applications". Data Mining and Knowledge Discovery, 5, pp 115-153, 2001

- [2] P. Giudici and R. Castelo, "Association models for web mining", Data Mining and Knowledge Discovery, 5, pp 183-196, 2001
- [3] B. A. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and evaluation of aggregate usage profiles for web personalization", Data Mining and Knowledge Discovery, 6, pp 61-82, 2002
- [4] W. Lim, S. Alvarez, and C. Ruiz, "Efficient adaptive-support association rule mining for recommender systems", Data Mining and Knowledge Discovery, 6, pp 83-105, 2002
- [5] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan, "Web usage mining : Discovery and applications of usage patterns from web data", SIGKDD Explorations, ACM SIGKDD, Vol 1, Issue 2, pp 12-22, Jan. 2000
- [6] Web Analyzer-Web Analytics Ver 3.0, ECMiner
- [7] 마이크로소프트웨어 월간지, 5월, 2000
- [8] <http://www.pionsoft.com/>, PIONSOFT