

# 웹 로그 분석에 기반한 상품 추천기의 설계

김건량\* · 이도현\*

\*전남대학교

## Design of a Product Recommender based on Web Log Analysis

Geon-lyang Kim\* · Doheon Lee\*

\*Chonnam National University

E-mail : glkim@dblab.chonnam.ac.kr

### 요 약

전자상거래의 활성화로 인터넷상에는 많은 쇼핑몰이 존재하고 정보도 또한 광범위하게 되었다. 그에 따라 고객이 원하는 정보를 얻기까지 소요되는 시간과 노력을 절약하기 위해 고객을 위한 상품 추천 시스템의 필요성이 증가하고 있다. 본 논문에서는 웹 로그 파일에 데이터 마이닝 기법을 이용하고, 고객의 행동 패턴과 그밖에도 고객의 프로파일, 상품 구매 데이터 등을 분석한 결과들을 기반으로 구축한 상품 추천 시스템을 제안한다. 추천 시스템은 고객에게 메일이나 우편을 보내거나 고객이 쇼핑몰에 방문하였을 때 웹 페이지를 추천해 주는 기능을 수행하여 고객이 원하는 정보를 쉽게 얻을 수 있도록 하는 편리함을 제공한다.

### ABSTRACT

As a lot of people have used electronic commerce, many shopping malls have appeared on the Internet and the shopping information in them has been enormous. So, the need for a system to recommend product to customers is on the increase so as to reduce time and efforts for shopping. In this paper, we suppose a Product Recommender System which is constructed by applying data mining techniques to web log files and analyzing customer's action pattern, customer's profile and product purchase data. This system offers convenience that customers can get their desired information easily, by sending e-mail or mail and recommending web pages when they visit a shopping mall.

### 1. 서 론

최근 인터넷은 정보의 홍수라 할 만큼 많은 정보들로 넘치게 되었다. 기존의 웹사이트에서는 단순한 정보 제공에 중점을 두었지만, 웹사이트의 콘텐츠 량의 증가에 따라 웹 페이지의 수도 증가하게 되었고, 그로 인해 고객들이 정보를 얻고자 하였을 때 인터넷은 많은 부담을 주게 되었다.

인터넷의 이러한 광대한 정보로 인해 고객은 원하는 정보를 쉽게 찾지 못하고 한번 방문한 웹사이트는 재 방문 확률이 매우 낮기 때문에 업체 측은 고객들을 놓치기 쉬웠다. 그러므로 고객들이 상품을 구매하도록 지원하고, 구매 고객을 확보하기 위한 비즈니스 전략이 필요하게 되었다. 고객을 유혹, 유치시킨 후 지속적인 관계를 유지시킬 수 있는 새로운 개념의 전략이 바로 개인화(Personalization)이다. 개인화는 고객의 개인적 취향에 따라 자신의 페이지를 구성하거나, 고객의

구매기록, 취향에 맞는 제품을 추천 받을 수 있는 기능들을 제공하는 등 차별화된 콘텐츠 제공을 통하여 고객에 대한 서비스를 극대화한 것이다.

이런 개인화 전략이 화두로 등장하면서 이에 대한 해결책으로 추천 시스템이 제시되기 시작하였다. 추천 시스템이란 기업이 제공하고자 하는 상품이나 정보 중에서 개별 고객이 선호할 것으로 예상되는 것들을 자동으로 찾아내 고객에게 제시해주는 개인화를 위한 지능적인 소프트웨어 시스템이다.

추천 시스템은 해외 업체의 넷퍼셉션(Netperception)이 대표적이며, 그 밖에도 브로드비전(Broadvision), 오토노미(Autonomy) 등이 있고, 국내에서는 (주)아이웍스가 최초로 개발과 상용화에 성공한데 이어 지속적인 개발 중에 있다. 추천 서비스를 제공해 주는 사이트는 해외의 Amazon, CDNOW를 비롯하여 국내의 Music OK, Shopping OK 등이 있다[1][2]. 이와 같이 많

은 추천 시스템이 개발되고 있으나 아직도 고객의 성향을 잘 파악하고 고객이 진정으로 원하는 상품을 추천해주기 위한 개발은 계속적으로 필요하다.

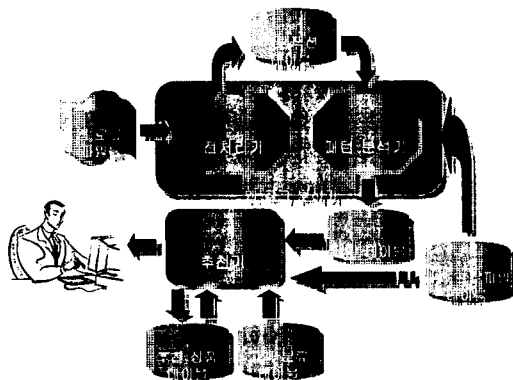
본 논문에서는 개인화 기술과 웹 로그 분석을 결합한 추천 시스템을 제안한다. 본 논문이 제안하는 추천 시스템은 고객의 성향과 비슷한 고객 그룹의 구매 패턴을 분석하고 다른 고객의 상품에 대한 만족도를 측정한다. 기존의 시스템과 달리 웹 로그 분석을 추가하여 앞으로의 구체적인 고객의 구매 성향을 파악할 수 있으며 우수 고객을 추출하여 우편이나 메일을 발송하기에 광고 비용을 절감할 수 있는 이점이 있다.

본 논문의 구성은 다음과 같다. 2장에서 시스템의 전체적인 구조를 살펴보고, 3장에서는 추천기의 구조를 살펴보고, 마지막 5장에서 결론으로 마무리하고자 한다.

## II. 상품 추천 시스템의 전체적인 구조

시스템은 웹 로그 파일을 입력 값으로 하여 전처리를 한 후에 관리자의 요구가 반영된 각 파라미터에 따라 작업을 행하고 해당하는 데이터를 트랜잭션 테이블(Transaction table)에 저장한다. 패턴 분석기(Pattern Analyzer)는 트랜잭션 테이블의 데이터와 고객의 신상 정보가 저장된 고객 테이블(Customer Table)의 데이터와 함께 마이닝 기법을 적용하여 패턴을 분석한 후에 패턴 테이블(Pattern Table)에 저장한다.

패턴 테이블에 저장된 지식은 메일이나 우편, 사이트에 방문한 고객에게 웹 페이지를 추천해주는 등 고객 지원을 위해 쓰이게 된다. 이러한 전체적인 웹 로그 분석을 위한 작업은 관리자가 일정한 시간을 정하여 오프라인에서 주기적으로 일괄 처리한다.



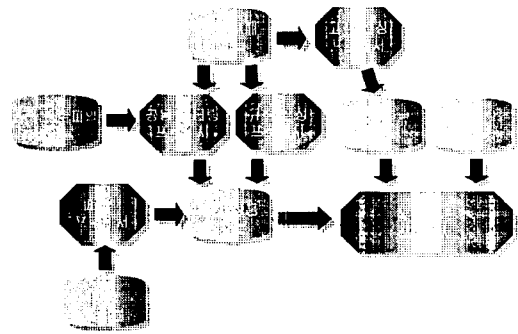
[그림 1] 시스템의 전체 구조

추천기는 고객의 상품 구매 테이블(Product Purchase Table)과 고객 프로파일 테이블(Customer Profile Table) 등의 데이터를 토대로 고객에게 상품을 추천해 준다. 시스템의 전체적인 구조는 [그림 1]과 같다.

## III. 추천기의 구조

본 논문이 제안하는 추천기는 웹 로그 데이터, 고객의 프로파일 데이터, 고객의 상품 구매 데이터, 상품에 대한 만족도 등의 정보를 기반으로 고객이 쇼핑물에 접근하였을 때 관심이 있어 구매 가능성이 있는 상품들을 추천 페이지로 제공하거나 메일, 우편 등을 발송한다.

추천기는 웹 로그 기반 프로세서, 공동 필터링 프로세서, 규칙 생성 프로세서, 피드백 프로세서, 고객 선정 프로세서로 구성된다. 웹 로그 기반 프로세서는 웹 로그 파일의 데이터에 데이터 마이닝 기법을 적용하여 유용한 규칙들을 생성하고, 공동 필터링 프로세서는 고객과 비슷한 성향을 가진 그룹의 선호도를 파악하여 상품을 추천한다. 규칙 생성 프로세서는 고객의 구매 데이터에 데이터 마이닝 기법을 적용하여 유용한 규칙을 생성하며, 피드백 프로세서는 고객에게 구매 상품에 대한 만족도를 측정하며, 고객 선정 프로세서는 특정 기준에 따라 우수 고객을 추출한다. 추천기의 구체적인 구조는 [그림 2]와 같다.



[그림 2] 추천기의 세부 구조

### 1. 웹 로그 기반 프로세서

웹 로그 기반 프로세서는 웹 로그에서 고객의 행동 패턴을 발견한다. 고객이 아직 상품을 구매하지는 않았지만 많이 보았던 상품 또는 고객과 비슷한 성향을 가진 고객들이 관심있는 상품은 구매할 가능성이 크다. 고객이 구매한 상품 데이터를 이용하여 분석하는 방법도 고객의 원하는 상품을 분석하는 데에 완전한 방법은 아니다. 고

객의 금전적인 사정 또는 다른 쇼핑몰에서 구매한 경우 등은 고객이 관심이 있어서 여러 번 상품 정보를 보았을지라도 구매 데이터에 기록이 되지 않으므로 구매 데이터를 기반으로 분석한 방법으로는 고객의 구매 욕구를 올바르게 파악할 수 없다. 웹 로그 분석은 고객이 방문했던 모든 상품 소개 페이지를 기록하기 때문에 이러한 단점을 보완하여 고객들의 구매 성향을 파악할 수 있으며 또한 고객들의 방문을 구매로 유도할 수 있다.

그런데, 웹 로그 파일에는 고객의 행동 패턴 분석을 위한 데이터 외에 많은 노이즈가 존재한다. 또한 웹 로그 데이터에 데이터 마이닝 기법을 적용하기 위해서는 웹 로그 데이터들을 적절한 형태로 가공하는 전처리 작업이 필요하다.

전처리를 위한 작업 순서는 다음과 같다. 먼저 데이터 추출 프로세서는 웹 로그 파일에서 그림 파일 등의 필요 없는 데이터는 삭제하고 마이닝을 적용할 대상이 될 데이터만 추출하고, 사용자 식별 프로세서는 각 사용자의 데이터들을 식별해 낸다. 다음 세션 식별 프로세서는 각 사용자들의 정보를 세션으로 식별해 내고, 패스 컴플리션은 한 세션에서 클라이언트의 브라우저와 프록시 서버의 캐시로 인해 로그 파일의 중간중간에 기록되지 않은 웹 페이지들을 채워주며, 트랜잭션 식별 프로세서는 마이닝 기법을 적용하기에 알맞은 트랜잭션 단위를 식별해 낸다[3].

트랜잭션은 접근 페이지들간의 연관 규칙을 얻기 위해 필요한 접근 페이지들의 모음으로 트랜잭션의 범위에 따라 또는 트랜잭션을 구성하는 항목들에 따라 얻어지는 연관 규칙의 의미가 달라진다. 본 논문이 제안하는 전처리기의 트랜잭션 식별 프로세서는 관리자의 요구에 따라 각기 다른 트랜잭션 테이블을 구축할 수 있다. 관리자가 원하는 트랜잭션 테이블을 구축하기 위해 시스템이 제공하는 선택적인 작업은 다음과 같다.

첫 번째, 고객에게 페이지를 추천할 때 모든 고객 또는 특정 고객 집합의 행동 패턴만을 고려하여 페이지를 찾을 수도 있기 때문에 고객에 따른 분석을 생각해 볼 수 있다.

두 번째, 페이지의 속성에 따른 트랜잭션을 구성할 수 있다. 고객은 보통 상품을 소개한 페이지를 원하는데, 상품을 소개한 페이지를 제외한 다른 페이지들은 상품을 소개한 페이지를 가기 위해 거치는 페이지이다. 그러므로 관리자의 분석 방법에 따라, 모든 페이지에 대해서 마이닝을 적용할 것인가? 아니면 상품을 소개한 페이지에 대해서만 마이닝을 적용할 것인가?를 구분하여 처리할 수 있다.

세 번째, 관리자가 지정한 기간 단위로 트랜잭션을 구성할 수 있다. 보통은 일괄 처리 기간 동안 각 고객에 대해서 성립하는 사건간의 연관 규칙만을 고려한다. 그러나, 어떤 연관 규칙은 전체 기간 동안 각 고객에 대해서는 신뢰도가 높지 않지만, 특정 기간 동안 특별히 강한 신뢰도를 가질

수 있다.

고객에 따라, 페이지의 속성에 따라, 시간에 따라 각기 달리 구성된 트랜잭션에서 고객들에게 페이지를 추천하기 위해 원하는 연관 규칙을 다양하게 얻을 수 있다.

## 2. 공동 필터링 프로세서

공동 필터링 프로세서는 유사한 취향을 가진 고객들이 공통적으로 선호할 것이라 예상되는 정보를 추천하는 방식이다. 이를 위해 비슷한 취향을 가진 고객들을 그룹화하는 작업이 필요하다. 본 논문이 제안하는 시스템의 공동 필터링 프로세서는 고객의 프로파일을 참조하여 취미별, 성별, 직업별, 연령별, 관심 있는 디렉토리별로 고객들을 그룹화하고, 각 그룹의 고객들이 주로 구매했던 상품들을 필터링하여 고객에게 추천한다. 또한 특정 상품을 구매하고자 할 때 고객에게 그 상품을 구매했던 고객 그룹이 많이 구매한 상품을 추천한다. 만약 음반을 구매하는 경우 그 음반을 구매했던 고객들이 많이 구매한 또 다른 음반뿐만 아니라 그 음반의 가수를 선호하는 고객들이 주로 구매하는 또 다른 음반을 추천할 수 있다. 공동 필터링 프로세서를 통해 고객과 유사한 취향을 가진 다른 고객 그룹의 행동 패턴을 반영할 수 있으므로 구매한 상품 정보들을 통해 고객의 구매 욕구를 촉진한다.

## 3. 규칙 생성 프로세서

연관 규칙 생성 프로세서는 트랜잭션 집합에서 연관 규칙을 도출한다. 트랜잭션 집합은 고객이 구매한 상품 데이터들이 특정 기간 단위로 구성된 기간 단위별 트랜잭션 집합, 특정 고객 그룹의 고객들이 구매한 상품 데이터들로만 구성된 고객별 트랜잭션 집합, 특정한 기간동안 고객이 구매한 상품 데이터들로 구성된 기간별 트랜잭션 집합이 있다.

기간 단위별 트랜잭션 집합은 고객이 구매한 상품 정보들을 하루나 이틀 등의 기간 단위로 트랜잭션을 구성하는 것이다. 일주일 단위로 트랜잭션을 구성한다면 일주일 동안 고객이 구매한 상품 데이터들이 하나의 트랜잭션이 된다. 이러한 트랜잭션으로 구성된 집합에서는 고객이 컴퓨터를 구입하면 일주일 내에 컴퓨터 책상을 구입한다와 같은 연관 규칙을 얻을 수 있다.

고객별 트랜잭션 집합은 고객이 구매한 상품 데이터들이 하나의 트랜잭션이 되고, 취미별, 직업별, 연령별, 성별로 고객들을 그룹지어 동일한 속성을 가진 고객들이 구매한 상품 데이터들이 트랜잭션 집합을 이룬다. 이러한 트랜잭션 집합에서 직업이 학생인 고객들의 50%는 만년필과 잉크를 구매한다와 같은 연관 규칙을 생성한다.

기간별 트랜잭션 집합은 계절별, 달별과 같이 특정한 기간 동안 고객들이 구매한 상품 정보들로 구성된다. 기간별 트랜잭션 집합은 여름이나 8월 등 특정 기간에 고객들이 어떻게 구매 성향을 보이는지를 분석한다. 예를 들어 여름은 수영복이나 양산 등이 잘 팔릴 수 있고, 8월의 경우 추석 명절이 포함되어 있어 제사 음식이나 가족들의 선물에 해당되는 상품들이 잘 팔린다는 등의 규칙을 생성할 수 있다.

#### 4. 피드백 프로세서

피드백 프로세서는 고객들에게 상품 구입 후 만족도를 증가시키기 위해 피드백을 제공한다. 고객이 상품을 구매한 후 상품에 대해 '아주 좋다', '좋다', '보통이다', '나쁘다', '아주 나쁘다'의 다섯 레벨이나 별의 개수로써 평가를 하도록 함으로써 다른 고객들이 상품을 구입하기 전에 좋은 참고가 되며, 상품을 구매한 후의 만족도 또한 증가하게 된다. 이러한 만족도는 상품 테이블에 저장되어 추천기가 메일이나 우편 발송 시 또는 추천 페이지 구성 시에 후보 추천 상품들에서 고객의 만족도를 높여주는 추천 상품들을 다시 필터링하는 역할을 한다.

#### 5. 고객 선정 프로세서

고객 선정 프로세서는 고객의 최근 구매 정보, 구매 빈도, 구매 금액으로 우수 고객들을 추출하여 메일이나 우편 대상 고객으로 선정하는 것이다. 예를 들어 최근 6개월 이내 구매한 고객, 신규 가입한 후 지난 월수로 구매 빈도를 나누어 산출된 빈도 계수가 일정치 이상인 고객, 구매 금액이 20만원 이상인 고객들을 우수 고객으로 선정하여 메일이나 우편을 발송하는 것이다. 모든 고객들에게 광고를 해야 하므로 많은 노력과 경비가 들었던 기존의 단점을 보완하여 본 시스템은 고객 선정 프로세서를 통해 우수 고객들만을 선정하여 상품을 추천하므로 광고비용을 절감하게 된다.

### IV. 결 론

본 논문에서는 인터넷 쇼핑몰에서 고객들을 지원하기 위해 웹 로그 분석을 이용한 상품 추천 시스템을 설계하였다.

전처리 과정에서 기존의 시스템은 데이터베이스에서 트랜잭션을 동일하게 구분했던 것에 반해, 본 논문에서 제안한 시스템은 원하는 파라미터들을 선택하여 관리자의 요구에 따라 각각의 트랜잭션을 구축함으로써 상황에 맞는 연관 규칙을

획득할 수 있다.

또한 추천기에서는 고객의 프로파일, 상품 구매 데이터, 고객의 상품에 대한 만족도, 그리고 웹 로그 분석 결과를 이용하여 고객의 구매 성향을 더욱 더 정확히 파악함으로써 고객들에게 만족된 개인화 서비스를 할 수 있었다.

### 참고문헌

- [1] J.B. Schafer, J. Konstan, and J. Riedl, "Electronic Commerce Recommender Applications", *Journal of Data Mining and Knowledge Discovery*, January, 2001.
- [2] J.B. Schafer, J. Konstan, and J. Riedl, "Recommender Systems in E-Commerce", In *Proceedings of the ACM Conference on Electronic Commerce*, November, 1999.
- [3] B. Mobasher, R. Cooley, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", *Journal of Knowledge and Information Systems*, Vol. 1, No. 1, 1999
- [4] B. Mobasher, R. Cooley, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", In *Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence*, November 1997.