

eCRM을 위한 웹 데이터마이닝 엔진의 설계 및 구현

○
김남호

호남대학교 정보통신공학부

Design and Implementation of Web Datamining Engine for eCRM

○
Nam-Ho Kim

Dep. of Information & Communication, Honam University

E-mail : nhkim@honam.ac.kr

요 약

전자상거래의 보편화와 더불어 인터넷 쇼핑몰이 급격히 늘어나면서 선택의 폭이 커짐과 동시에 고객에게는 선택의 불편이 있고, 판매자에게는 경쟁이 가중될 것이므로 더욱 치밀하고 편리한 마케팅 기법이 요구되어지고 있다. 따라서 인터넷 쇼핑몰에서의 개별고객의 니즈를 파악하기 위한 다양한 분석적 접근이 이루어지고 있다. 웹 데이터마이닝은 웹사이트에 접근한 사용자들의 개인별 접근패턴을 파악하고 예측하기 위해 로그파일을 분석한다. 본 연구에서는 eCRM의 핵심요소인 로그분석을 위한 데이터마이닝 엔진을 설계하고 구현한다.

1. 서론

인터넷쇼핑몰의 활성화에 따라 고객이 접하는 정보 또한 급격히 증가하게 되었다. 고객의 입장에서 선택의 폭이 넓어지게 되었지만, 수많은 정보 안에서 고객이 진정으로 원하는 정보를 얻기 위해 시간과 노력이 많이 필요하게 되었고, 업체의 입장에서는 정보제공자들의 증가로 경쟁이 치열해지게 되었다. 고객관계관리(eCRM, Customer Relationship Management)로 대표되는 고객중심적 마케팅으로의 패러다임 변화가 일고 있는 것이다.[5] 특히 인터넷은 회사의 상품과 서비스를 마케팅하기 위한 강력한 매개체로 등장하였다.

인터넷마케팅은 가상의 공간에서 소비자와의 관계 형성 및 실시간 상호작용이 가능한 쌍방향 커뮤니케

이션을 통한 일대일 개별화 마케팅이 실현될 수 있는 최적의 환경을 제공하고 있다. 즉, 개별 고객의 특수한 니즈를 개별적으로 파악해서 각각의 고객에게 차별화된 서비스를 제공하는 것이 그 핵심이다. 이때 기업은 개별 고객과의 관계증진을 통해 고객 유지율을 상승시키고 해당 회사에 대한 충성심을 유도하여 결국은 이익을 증대시킬 수 있는 것이다.

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

로그파일 분석의 데이터마이닝 결과는 eCRM 주제 중 하나인 상품추천 및 교차판매(Cross Selling)에 적용된다.

본 연구에서는 로그파일을 분석하기 위한 데이터 모델링 및 데이터마이닝 엔진을 설계하고 구현한다. 또한 결과로서 고객관계관리(eCRM)에 미치는 효과를 분석하였다.

논문구성은 2장의 관련연구, 3장에서는 웹 데이터마이닝 엔진 설계 및 구현에 대해서, 마지막 4장 결론에서는 기대효과 및 향후 연구에 대하여 논하였다.

2. eCRM과 데이터마이닝

eCRM과의 통합을 통해 인터넷비즈니스는 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.

효율적인 eCRM을 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 최초에 입력되는 고객 정보는 신뢰성이 부족하여, 지속적으로 갱신되지 않으면 eCRM을 통해 얻을 수 있는 이점은 거의 없다.

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

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 데이터마이닝을 통해서 온라인 상에서 제공하는 서비스와 제품간의 연관관계를 밝혀내어, 적절한 제품이 적절한 서비스와 함께 판매

가 되고 있는지의 여부를 밝혀내게 될 수도 있다.

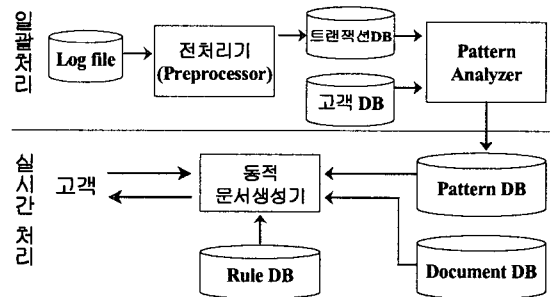
데이터마이닝을 통해서 기업은 웹사이트상의 패턴을 의미 있는 정보로 종합해내고, 인터넷상의 고객들과 예상치들을 이해하고 연관시킬 수 있게 된다. 데이터와 웹이 제공하는 방대한 사업지식의 흐름에 근거한 웹 데이터마이닝은 온라인 고객과의 관계를 생성하고 유지시키며 생산성 있는 온라인 상점의 최전선을 구축하는데 있어 결정적 열쇠가 되는 것이다.

본 논문에서 제안한 데이터마이닝 엔진은 개인별 서비스 제공을 목적으로 한다. 고객이 웹사이트에 방문하여 등록할 때 입력한 고객의 신상정보와 고객이 접근한 시간, IP주소, Agent 등의 정보를 바탕으로 고객이 재 방문하였을 때 고객이 관심 있을만한 정보를 동적으로 서비스해 준다.

3. 웹 데이터마이닝 엔진의 설계 및 구현

1) 시스템 구조

제안한 엔진은 [그림1]과 같이 크게 전처리기(preprocessor), 패턴분석기(pattern analyzer), 문서생성기(document organizer)의 세 모듈로 구성된다.



[그림1] 시스템엔진의 구조

시스템은 사용자들의 접근 기록을 가지고 있는 로그파일을 입력으로 패턴분석이 가능한 형태로 로그 기록을 포맷하는 전처리기와 포맷된 로그 기록의 분석을 담당하는 패턴분석기로 구성되어있으며, 이 두 모듈은 일정기간에 한번씩 실행되는 일괄처리 기법을 사용한다. 그리고, 분석된 결과에 의해 사용자 개인별 동적 문서를 생성하는 문서생성기는 실시간 처리를 기본으로 한다.

2) 전처리기

일반적으로 전처리기는 로그파일에서 필요한 정보만을 추출하는 데이터클리닝(data cleaning), 사용자별로 로그기록을 분류하는 사용자분류(user identification), 개인별 기록을 세션으로 분류하는 세션분류(session identification), 생략된 로그기록을 완성시키는 패스 컴플리션(path completion), 그리고 패턴분석기의 입력으로 사용될 트랜잭션을 생성하는 트랜잭션 분류(transaction identification)의 다섯가지 서브모듈로 구분된다.[5]

데이터클리닝이란 웹로그 분석을 위해 부적합하다고 생각되는 항목들을 웹로그 파일에서 제거하는 것이다. 기존의 시스템이 웹서버의 로그파일을 이용하여 유용한 정보만을 얻는 데이터클리닝이었던 반면, 제안한 시스템에서는 사용자의 웹서버 접근시 사용자의 IP주소와 접근시간, Agent등의 정보가 실시간으로 데이터베이스에 저장되므로 따로 데이터 클리닝을 할 필요가 없다.

사용자가 로그인 후 웹서핑을 하게되면 사용자의 정보는 사용자 로그 데이터베이스에 저장되고, 기존 고객이 웹서핑을 하다가 로그 온을 하거나 첫 고객이나중에 등록을 할 경우는 로그 데이터베이스에 저장된다. 그리고 로그 데이터베이스에 저장된 정보는 로그 온을 한 후에 고객ID와 IP주소, 시간, Agent가 저장된 접근 히스토리 데이터베이스를 참고하여 사용자 로그 데이터베이스에 다시 저장된다.

사용자 분류기는 사용자 로그 데이터베이스에 저장된 레코드들을 사용자별로 분류한다. 이때 고려할 점은 IP주소가 같아도 다른 브라우저이거나 다른 운영체제이면 다른 사용자일 수 있다는 것이다.

세션분류기는 사용자 로그 데이터베이스에 있는 사용자의 웹페이지 접근을 개별적인 세션으로 분류한다. 세션을 분류할 때 보통 타임아웃 방법을 사용하는데, 제한시간을 초과하였을 때 사용자가 새로운 세션을 시작한다고 가정한다.

기존의 로그파일 분석에서는 사용자가 백 버튼을 선택했을 때 프록시 서버나 쿠키를 사용하게되면 웹서버 로그파일에 기록이 남지 않아 패스 컴플리션 과정이 필요했다. 하지만 여기서는 사용자 접근 정보가 데이터베이스에 기록이 됨으로 이러한 과정이 필요하지 않다.

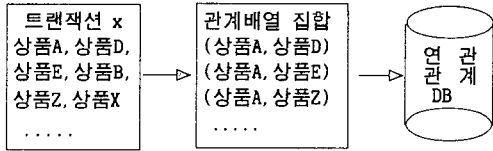
트랜잭션 분류기는 각 사용자에 대해 의미있는 관계

들의 집합을 생성하기 위해 최대 시간 간격으로 세션을 여러 트랜잭션으로 분류한다.

3) 패턴 분석기

데이터마이닝 모듈인 패턴분석기에는 전처리기의 트랜잭션에 포함된 접근한 웹페이지가 항목으로 입력된다. 이 웹 페이지에 연관규칙(Association Rule)을 적용하여 접근한 웹 페이지들간의 패턴지식을 찾아내고 패턴 데이터베이스에 저장한다.

전처리과정을 통해 정제된 데이터는 다음과 같은 방법으로 분석하여 페이지 항목간 연관관계를 계산한다. 쇼핑몰사이트의 경우를 예로 들면, 상품데이터의 상품 코드로 로그데이터의 한 트랜잭션 내에서 X라는 제품을 액세스한 고객이 Y라는 제품을 액세스한 연관관계를 검사한다. 각 트랜잭션단위에서 한 고객이 액세스한 상품들간의 관계를 관계배열 집합으로 생성하여 상품 연관관계 테이블에 저장한다. 최종적으로 상품간의 연관관계 지도도와 신뢰도를 산정한다.



[그림2] 연관관계 분석

4) 문서 생성기

고객이 쇼핑몰에 접근하였을 때 패턴 데이터베이스에 저장된 규칙을 적용하여 고객의 성향에 맞는 웹페이지를 동적으로 구성하여 제공한다. 예를 들어 교차판매 방식에 의한 상품 프로모션 과정을 설명하면 다음과 같다.

- 프로모션 대상 지지도 결정
- 상품간 교차판매 대상 확인
- 고객이 상품보기를 하였을 때 관련 상품리스트로 웹사이트에 프로모션
- 관련상품을 웹페이지에 리스팅할 때 연관관계에 의해 교차판매임을 명시하는 코드를 추가하여 실제 페이지뷰가 발생했을 때 로그에 기록

이때 이용되는 규칙 데이터베이스는 글자의 크기와 모양, 선 등 문서의 전체적인 디자인을 결정해 주는 요소를 포함하고 있다. 이 문서는 각 고객의 구매패턴, 관심분야에 따라 각기 다른 링크정보와 우선 순

위, 색깔, 크기 등을 갖는다.

5) 시스템 평가

활용성을 중심으로 제안한 시스템을 평가하고자 한다. 웹사이트를 통한 소비자 구매 행동의 추적 및 이를 활용한 eCRM시스템 구축은 현 수준에서는 아마 남의 생각을 읽는 독심술 이상으로 막강한 마케팅 무기로 활용할 수 있을 것이다. 인터넷 사이트 운영자와 쇼핑객들이 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.

접근행동 활용을 통한 타겟 마케팅이란 어떤 것인지 제안한 시스템의 예를 들어 살펴보자. 기념일을 맞아 쇼핑객의 꽃배달 서비스를 통해 장미패키지를 구매한 사람은 동일한 연인에게 보낼 샐 샐 향수를 화장품코너에서 구매하고, 또 뮤직코너에서 좋아하는 가수의 CD를 구매했다고 한다. 이 사람에게는 향후 구매했던 아이템을 근거로 일정 형태의 구매 패턴을 갖고있는 것이 확인되므로 그에 맞는 제안형 추천을 함으로써 직접적인 매출을 유발할 수 있을 것이다.

4. 결론

연관규칙을 이용한 데이터마이닝 기술을 사용자의 웹 페이지 접근 데이터 분석에 적용하기 위한 데이터 처리 시스템과 생성된 규칙을 이용한 마케팅 활용 기법들을 소개하였다.

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 중요시되는 현 시점에서 쇼핑객이나 웹사이트 운영 기업의 고객관계관리(eCRM)를 위한 중요한 방안이 될 것이다.

본 논문에서 제안한 데이터마이닝 엔진은 쇼핑객을 운영하는 기업의 입장에서 지속적인 고객관리에 목표를 둔 eCRM 시스템의 핵심 요소이다. 아울러 지속적인 증가추세에 있는 모바일 환경에 맞는 비즈니스 모델 개발과 여기에 맞는 CRM에 관한 확장연구가 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, 4th World Congress on Expert system 98, March 16-20, Mexico City.
- [2]Bamshad Mobasher, Robert Cooley and Jaideep Srivastava, "Automatic Personalization based on Web Usage Mining", Communications of the ACM, 43(8), pp.142-151, 2000.
- [3]한국오라클, "Oracle CRM: Release11i Statement of Direction ", 오라클 백서, 2000.
- [4]한재홍, "CRM특강", <http://www.dpc.or.kr/dbworld/document/9912/tech-1.html>
- [5]Bamshad Mobasher, Cooley, J. Srivastava, "Data Preperation for Mining World Wide Web Browsing Patte군", in Journal of Knowledge and Information System, Vol. 1, No. 1, 1999.
- [6]H. Edelstein, "Building Profitable Customer Relationships with Data Mining", SPSS White Paper, 2000.