

eCRM을 위한 데이터마이닝 에이전트의 개발

(The Development of the Data Mining Agent for eCRM)

손달호*, 홍덕훈*

(Dal-Ho Son, Duck-Hoon Hong)

요약 전자상거래의 성장에 따라 고객의 특성을 파악하고 고객의 기호에 맞는 정보를 제공하여 구매력에 영향을 주고 동시에 웹사이트 운영자의 정보선별과 관련된 의사결정을 지원하는 시스템도구들이 개발되고 있다. 그러나 이런 시스템도구들은 웹운영기업에 경제적 부담을 주고 컴포넌트화 되어 있지 않아서 원하는 데이터를 수집하기보다는 수집된 데이터에 기업들의 요구정보를 맞춰야 하는 어려움이 있다. 이러한 점을 감안하여 본 연구는 기업의 필요정보에 대한 맞춤형 웹에이전트 시스템의 설계로 도구기반 로그분석을 대신할 수 있는 시스템개발을 연구목적으로 설정하였다. 본 연구에서 개발된 웹에이전트는 설계언어로 Java와 ASP를 이용하였으며 데이터 저장방식은 쿠키를 이용하는 방식으로 구현되었고 장기적 보존 데이터를 위해서 My-SQL의 데이터베이스와 연동될 수 있도록 개발되었다. 본 연구에서 개발된 시스템은 데이터마이닝개념을 이용하여 도구에 분석데이터를 전달시키기보다는 적극적으로 웹사이트의 컴포넌트형 에이전트로 데이터를 가공하도록 하였고 따라서 고객 개인화 인터페이스 설계 뿐 아니라 의사결정과 전문가 시스템의 성격을 지닌 정보제공방식을 마련할 수 있다는 점에서 기여도가 있다고 본다.

핵심주제어 : 데이터마이닝 에이전트, 로그분석

Abstract Many attempts have been made to track the web usage patterns and provide suggestions that might help web operators get the information they need. These tracking mechanisms rely on mining web log files for usage patterns. The purpose of this study is to verify a web agent prototype that was built for mining web log files. The web agent for this paper was made by Java and ASP and the agent came into being as part of a cookie for a short-term data storage. For long-term data storage, the agent used a My-SQL as a Data Base. This agent system could inform that if the data comes from the web data mining agent, it could be a rapid information providing method rather than the case of data coming into a data mining tool. Therefore, the developed tool in this study will be helpful as a new kind of decision making system and expert system.

Key Words : eCRM, Data Mining Agent, Log Analysis

1. 서론

eCRM(Customer Relationship Management)시대 인터넷을 이용한 웹사이트는 마케팅 전략과 조직의 의사결정 지표설정을 위해 사용자 개개인 의 웹사이트 접근 방법에 대한 정보가 필요하다. 이와 같은 정보를 기반으로 웹사이트 운영자들은

사용자들의 취향에 맞는 동적인 개인화 인터페이스를 제공할 수 있게 된다.

이와 같은 목적을 위해 필요한 것이 웹로그 처리과정이며 로그처리에 이용되는 로그파일은 IP주소, 접근시간, 접근방법, 접근한 URL, 프로토콜, 에러코드등에 대한 내용을 포함할 수 있다. 이러한 로그파일의 분석을 통하여 주요고객층, 고객구매패턴, 구매시간, 구매탐색경로 등의 데이터를 추출할 수 있으며 추출된 데이터를 기반으로 인터페이스

* 계명대학교 경영정보학과

설계, 상품레이아웃 설계, 고객센터의 강화 등 다양한 대 고객마케팅을 수행할 수 있다. 최근 이와 같은 로그분석을 위해서 데이터를 수집하는 에이전트의 역할이 부각되고 있으며 전자상거래 보편화에 따라 다양한 모습의 온라인 에이전트가 등장하였으나 아직까지 그 기능을 완전히 소화해 내는 에이전트는 많지 않았다[6].

본 연구는 개인화 인터페이스를 제공하는 데이터마이닝 에이전트(data mining agent)의 개발에 대해 중점을 두고자 한다. 데이터마이닝 에이전트는 기존의 데이터마이닝툴(data mining tool)의 프로세스를 분석하여 개발된 것이다. 데이터마이닝툴은 웹에서 모아지는 데이터를 서버로 가져온 다음 일련번호에 따라 로그파일의 등록과 패턴분석의 탐색결과로 처리한다[4]. 또한 고객정보에 따른 동적 인터페이스를 제공하기 위하여 웹 자원이 재구성되는 형태로 로그파일을 간접적으로 사용하는 방법을 채택하고 있다. 이에 반해 데이터마이닝 에이전트는 서버에서 파일이 전달되기를 기다리지 않고 웹페이지의 구성에 컴포넌트적 요소로서 활동한다. 또한 로그파일의 생성을 동적으로 생성함으로써 인해 많은 정보를 확보할 수 있으며 기존 도구들에 비해 개발과 유지보수비용 측면에서 장점이 있다.

본 연구에서는 이와 같은 점을 감안하여 기존의 데이터마이닝툴과 다른 방법의 에이전트설계를 통하여 직접적인 로그파일의 사용으로 고객의 행동 패턴 분석과 그에 따른 개인화 인터페이스 제공가능성을 보여주고자 한다.

2. 이론적 배경

2.1 B2C로그분석

로그분석이란 웹관련 데이터를 기반으로 다양한 정보를 추출해 내는 것이라 할 수 있다. 로그분석은 사용자에 따라 단지 로그 정보를 분석하는 것에 한정시키기도 하고 로그 정보를 기반으로 다양한 정보를 분석하는 확장된 개념으로 확대시키기도 한다[7,10].

일반적으로 좁은 의미의 로그분석은 분석 대상

을 웹 로그에 한정 짓고 이에 대한 분석을 실시하는 것을 말한다[2]. 기본적인 페이지뷰부터 방문자의 방문경로, 방문자수 및 참조페이지 등을 측정할 항목들을 측정한다. 현재 주요 로그분석 도구들은 대체적으로 이와 같은 방법을 사용하고 있다.

최근에는 방문자 관계관리가 중요해지면서 방문자를 중심으로 로그데이터를 추출 및 관리하는 경향이 대두되고 있다[5]. 즉 로그인을 통해 접속한 사람과 그렇지 않은 사람에 대해 접근방법이 다른데, 전자의 경우는 사용자가 확인되었으므로 바로 사용자의 웹활용 현황이 파악 가능한 반면 후자의 경우는 쿠키를 사용하여 간접적으로 사용자를 확인한다.

이와 같이 확인된 사용자의 웹활용현황을 파악하여 IP기반으로 로그 데이터를 관리하게 되면 실질적으로 사이트에 주로 방문하는 사람은 누구이며 이들이 주로 관심을 가지고 보는 페이지는 무엇인지 파악할 수 있게 된다. 즉 개별 방문자와 관계를 맺고 일대일 마케팅을 수행해 나갈 수 있는 기반이 되는 것이다.

넓은 의미의 로그분석은 협의의 로그 분석에 반해 로그 데이터뿐 아니라 웹사이트에서 보유하고 있는 고객 등록 정보, 구매정보등을 분석대상으로 하는 것을 말한다[1]. 특히 거래정보를 기반으로 하는 고객분석은 데이터 충실도 측면에서 순수 로그분석 보다 우월하다는 특징이 있다. 또한 웹활용 DB가 고객등록 DB와 연동되는 경우 인구통계학적으로 웹활용현황의 특징을 파악할 수 있으며 사례기반추론기법 등을 활용하여 데이터마이닝 작업을 통해 콘텐츠 및 상품 등을 추천할 수 있다[6].

그러나 로그데이터를 분석하는 목적은 궁극적으로 방문객들이 웹사이트를 방문하면서 남긴 흔적들을 분석함으로써 고객들이 관심을 가지는 내용을 파악하고 고객들의 반응을 예측하여 인터넷 비즈니스에 전략적으로 활용할 수 있는 정보를 얻는 것이다. 따라서 로그분석툴을 도입하여 로그분석작업을 할 때 이와 같은 목적을 분명히 명시하지 않으면 분석작업은 분석으로만 끝날 수 있다.

로그분석을 통하여 웹사이트를 통해 이루어지는 모든 정보를 정확하게 파악할 수 있는 것은 아니다. 로그데이터 그 자체만으로 분석할 수 있는 정보는 그다지 많지 않은 것이 사실이다. 또한 로그

데이터 자체로는 완벽하지 않다는 특징을 가지고 있는 점도 인식하고 있어야 한다.

예를 들어 로그데이터에 저장되는 클라이언트IP를 살펴보면 사용자가 인터넷을 시작한 위치정보를 파악할 수 있다. 하지만 사용자가 유동IP를 사용할 경우, 또는 프락시서버를 이용할 경우 정확한 클라이언트IP를 파악할 수 없다. 또한 캐쉬를 통할 경우 사용자는 페이지를 보았지만 로그에는 데이터가 저장되지 않는다[14]. 그리고 “back키”를 통해 이동할 경우도 그 행위들은 다른도구를 이용하여 추적은 가능하지만 로그에는 기록되지 않는다.

이를 보완하기 위해 사용되는 쿠키의 경우에도 브라우저의 동일성을 파악할 수는 있지만 PC에서 두명 이상이 사용할 경우에는 정확한 사용자를 확인할 수 없다. 또한 클라이언트의 쿠키거부라든가 개인정보보호를 위한 쿠키삭제 프로그램에 대해서는 역시 대응 방안이 없다.

이러한 한계를 극복하기 위한 가장 좋은 방법은 온라인상의 고객정보수집을 단순 웹로그에 국한하지 않고 다양한 채널을 통해 고객 정보를 수집하여 통합하는 것이다. 가령, 콜센터, 게시판, 이메일, 핸드폰, 채팅 등 고객과의 온라인상의 접점은 웹에 국한되지 않도록 구성하도록 하는 것이다.

일반적인 데이터 수준의 로그만으로는 사용자 관점에서 유용한 정보가 되지 못한다. 일정한 정제 과정을 거친 후 산출된 데이터를 DB에 적재함으로써 좀더 유용한 자료가 될 수 있다. 이와 같이 로그데이터를 DB에 적재하기 전에 데이터를 전처리 하는 과정이 추출엔진에 의해 수행된다.

Cleaning 로그과정은 모든 데이터를 서버 스토리지로 가져갈 필요없도록 하기 위해 데이터를 제거하는 과정이다. 사용자확인과정에서는 사용자ID를 중심으로 확인과정에 활용하지만 고정아이피나 프락시서버를 활용하는 경우에는 IP주소를 활용하기도 한다. Session 확인과정에서는 로그인 시간을 통해 스트림 상태를 확인하는 단계를 의미한다. 본 연구에서 개발되는 시스템은 이와 같은 ID매칭기법을 이용하여 맞춤형 개인화기능을 제공할 수 있는 시스템이 될 것이다. 즉 맞춤형 개인화기능을 제공하기 위해 사용자들이 사용한 웹로그파일을 이용하여 좀더 사용자 중심적인 시스템으로 구축하고자 하는 것이 본 연구의 목적이다.

2.2 웹개인화

전통적 CRM과 eCRM의 목적이나 기대효과는 동일하지만 eCRM에서는 고객의 이해를 위해 수집하는 고객정보의 종류, 수집방법 및 고객에게 가치를 제공하는 채널이 인터넷을 기반으로 한다는 측면에서 차이가 있다. 또한, eCRM에서는 e-Business가 갖는 특성과 인터넷관련 기술을 이용하여 오프라인 CRM과는 차별화된 새로운 전략을 고려할 수 있으며, 차별화된 핵심전략 중 한가지 방법이 웹 개인화를 이용한 방법이다[8].

Amarasiri and Alahakoon[1]은 인터넷 비즈니스상의 모든 상품과 광고, 메뉴 등의 콘텐츠를 고객 개인의 성향과 취향에 맞도록 웹사이트를 개발하여 고객에게 필요한 정보를 제공하며 획일적인 서비스가 아닌 고객과 친근한 일대일 관계를 맺기 위한 웹 전략을 개인화라고 정의하였다. eCRM에서 고객 로열티 증대의 핵심기술인 개인화는 충분한 고객분석을 통해 고객관계의 현실적 개선에 공헌할 수 있는 주요가치를 어떻게 제공하는 가에 대한 전략이 될 수 있다.

개인화서비스는 상품을 선별하는 개인화 알고리즘이 충분히 지능화되어야 정확성을 보장할 수 있으며 또한 고객에 의한 상품들의 추천효과를 기대할 수 있다. 개인화를 구현하는 알고리즘으로는 아마존에 적용한 협업필터링(collaborative filtering)과 룰기반엔진(rule-based engine)이 있다[3]. 협업필터링은 어떤 고객의 성향과 비슷한 성향을 갖는 고객집단을 찾아 그와 유사한 집단의 패턴정보를 근거로 서비스를 제공하는 것이다. 룰기반엔진은 개인전략에 근거하여 마케팅 담당자가 직접 규칙을 미리 지정하여 자동으로 개인화가 수행되도록 하는 방식이다.

개인화관련 기술의 최근 추세는 협업필터링과 룰기반엔진을 통합하여 제공하고, 데이터마이닝의 기법인 연관분석 등을 이용하여 개인화 알고리즘을 통합적으로 제공하고 있다[5]. 또한, 개인화알고리즘 기술을 이용하여 추천 프로세스만을 적용하는 사이트에서는 데이터웨어하우스 중심의 비즈니스 인텔리전스를 추가 도입하여 고객에 대한 정확한 정보를 수집하고 있다.

Manber et al.[8]은 개인화 설계단계를 다음과

같이 나타내었다. 첫째 단계는 개인화 설계단계로 고객의 안전특성을 반영하는 단계이다. 여기에는 고객에 대한 기본정보인 성별, 연령, 직업, 주소, 취미 등이 포함된다. 둘째 단계는 고객과 회사가 접촉하고 거래하면서 드러난 고객의 접촉 및 거래 특성을 반영하는 단계로, 여기에는 구매상품의 유형, 구매주기, 웹페이지상의 관심정보 등이 포함된다. 셋째 단계는 고객의 인적특성과 접촉 및 거래 특성의 결합을 통하여 개인고객의 종합특성을 도출하는 단계이다. 개인고객의 종합특성은 하나의 변수로 요약할 수도 있고, 여러 개의 변수로 설명할 수도 있다. 마지막 단계로는 고객의 종합적 특성에 따라 적합한 오퍼를 제공하기 위한 개인화규칙을 결정해야 한다. 개인화 규칙이 결정되면 비로소 개인화된 오퍼를 제공할 수 있다. 개인화는 상품이나 부가적 혜택 모두에 적용 가능하나, 실제로 eCRM에서 많이 적용되는 개인화는 웹 콘텐츠를 개인화시키는 것과 고객에게 적합한 상품을 제안해 주는 것 등이다.

Riecken[9]은 개인화 설계단계에서의 핵심은 개인화규칙을 결정하는 것이라고 주장하였다. 개인화규칙은 고객의 특성과 제약조건을 이용하여 개인고객에게 적합한 내용을 추출하는 것이다. 개인화규칙은 고객의 특성과 오퍼를 연결시키는 역할을 하며, 따라서 개인화규칙은 eCRM에서 중요한 역할을 담당하고 있다. Amarasiri and Alahakoon[1]은 이와 같은 개인화유형을 다섯가지 유형으로 분류하였는데 인적특성에 기초한 개인화규칙, 선호하는 상품이나 정보에 기초한 개인화규칙, 결합조건을 고려한 개인화규칙, 제약조건에 기초한 개인화규칙 및 유사 사례에 기초한 개인화규칙 등으로 분류하였다. 본 연구에서 개발된 에이전트시스템은 이와 같은 개인화 유형들을 혼합하여 웹사용자 개인에게 좀더 맞춤형서비스를 제공할 수 있도록 구축함을 목적으로 한다.

3. 데이터마이닝 에이전트의 설계 및 구현

3.1 시스템 설계

<그림 1>에서 보는 바와 같이 기존의 데이터마

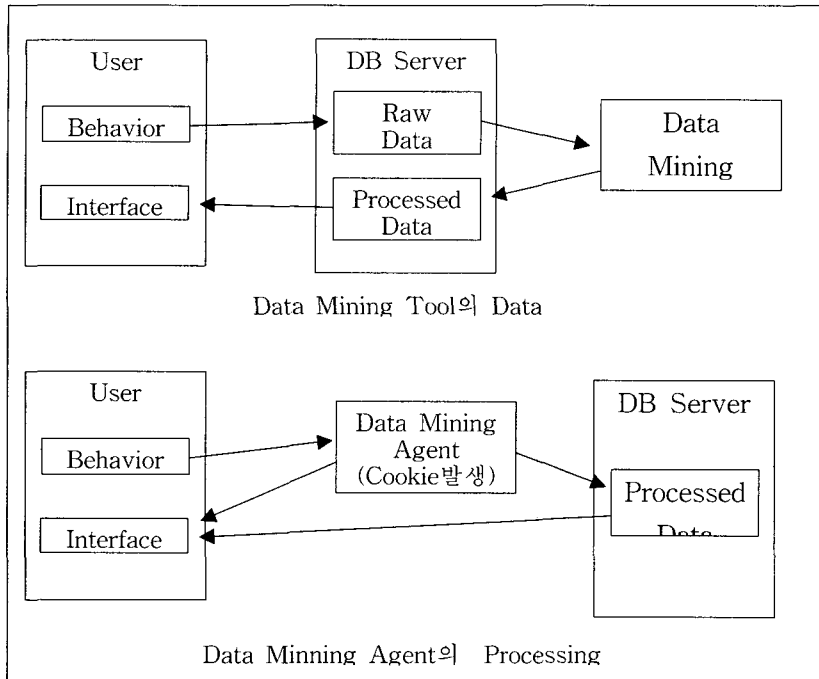
이닝들은 데이터를 서버로 가져와 처리자료로 재구성하였으나 데이터마이닝 에이전트는 비등록 기본키의 처리 능력을 포함하는 것으로 데이터가 서버에 전달되기 전에 처리한다. 본 연구에서는 데이터마이닝 에이전트 방법을 이용하여 데이터를 서버로 이동하지 않고 재구성하기 위해 쿠키를 사용하였다. 쿠키는 고객이 특정 홈페이지를 접속할 때 생성되는 정보를 담은 임시파일로 크기는 4KB 이하로 작으며, 초기에는 인터넷 사용자들의 홈페이지 접속을 돕기 위해 만들어졌다. 특정사이트를 처음 방문하면 아이디와 비밀번호를 대신한 쿠키가 만들어지고 다음에 접속했을 때 별도의 절차 없이 사이트에 빠르게 연결할 수 있다.

쿠키는 사용하는 웹브라우저가 자동으로 만들기도 하고 갱신하기도 하며, 웹사이트로 기록을 전달하기도 한다. 이용자가 인터넷에서 어떤 내용을 봤는지, 어떤 상품을 구입했는지 등 모든 정보가 기록되기 때문에 개인의 사생활을 침해할 소지가 있다. 온라인 광고업체들은 쿠키를 이용해서 인터넷 사용자의 기호 등을 수집·분석해 광고전략을 짜는 데 유용하게 활용해왔다. 또한 쿠키는 회원번호나 비밀번호 등이 유출할 가능성이 있기 때문에 보안문제를 유발할 수도 있다. 그래서 마이크로소프트는 인터넷 익스플로러 5.0 이상에서는 쿠키 거부 기능을 첨가하였다. 그러나 대다수의 콘텐츠제공 사이트가 쿠키를 거부하는 경우에는 서비스를 받을 수 없도록 웹사이트를 설정해 놓고 있으므로 본 연구에서 논의되고 있는 에이전트의 설계에도 이와 같은 보안문제를 고려하여 쿠키를 설계하였다.

본 연구에서 에이전트의 설계에 사용된 언어는 Java, ASP이며 데이터베이스는 MS사의 Access를 사용하였고 PWS와 MS-SQL을 통해서 연동되도록 설계하였다. 본 연구에서 제안한 에이전트는 하나의 사이트에서 고객을 관리하기 위해 두 개의 DB를 필요로 한다. 데이터베이스는 고객이 사이트 방문시에 로그인 했을 경우와 그렇지 않았을 경우로 나누어지는데 실제로 두 개의 DB를 사용하는 것은 시스템의 효율성과 데이터중복성 측면에서 좋지 않으므로 로그인을 통한 실제 거래가 이루어지는 데이터베이스는 웹 DB를 사용하고 그렇지 않은 경우는 쿠키를 통한 가상 DB를 사용하도록 설계하였다. 가상 DB는 고객의 데스크탑 컴퓨터를

데이터 저장대상으로 가상하지만 4KB의 쿠키에 텍스트와 숫자만 입력되므로 실제로 본 연구에서 개발된 에이전트에는 다수의 데이터가 저장된다고 볼 수 있다.

후 고객의 웹서핑데이터는 콤보버튼등을 클릭할 때 Onclick()명령으로 쿠키에 기록하게 되도록 설계되었다.



<그림 1> 데이터마이닝툴과 데이터마이닝 에이전트의 자료처리방법의 비교

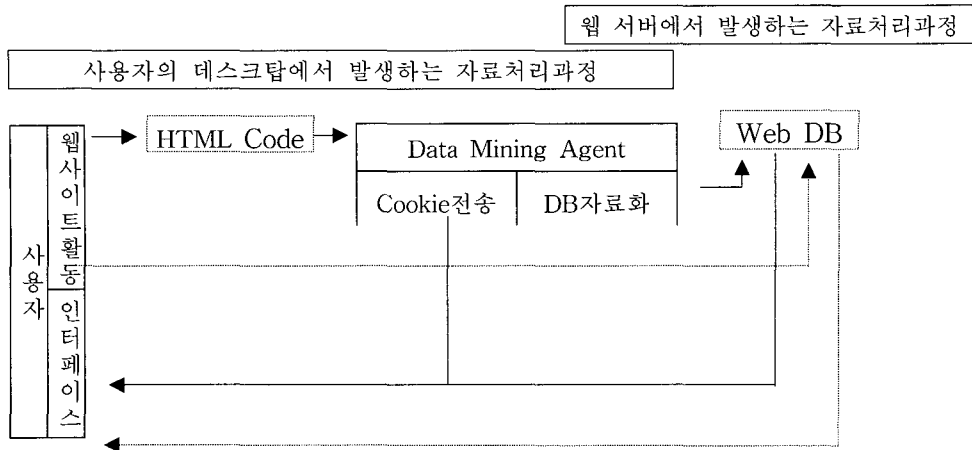
에이전트는 웹에서 사용자의 컴퓨터로 다운로드 하여 데이터 작업을 하므로 웹 DB로 옮겨가지 않는 데이터들은 사용자의 컴퓨터에서 즉시 생성 및 갱신되고 쿠키에 저장된다. 그 후 재접속시에는 사용자의 쿠키에 저장된 데이터와 그에 일치하는 데이터를 웹 DB에서 찾아내어 사용자의 화면에 표시한다. <그림 2>는 본 연구에서 설계된 데이터마이닝 에이전트의 이동경로를 보여주고 있다.

본 연구의 에이전트관련 프로그래밍에 대한 상세 설명을 플로우차트로 나타내면 <그림 3>과 같다. 사이트 최초 접근시에는 조건문을 만나서 쿠키의 존재 유무를 확인한다. 쿠키가 존재하는 경우에는 그 안에 기록된 값과 서버의 데이터중 일치하는 부분을 찾아 배너란에 고객을 위한 광고를 내보내고 쿠키가 발견되지 않았을 경우에는 쿠키를 사용자의 컴퓨터에 기록한다. 서버데이터를 화면에 표시하는 것은 사이트 접속시에 모두 일어나며 그

3.2 시스템구현

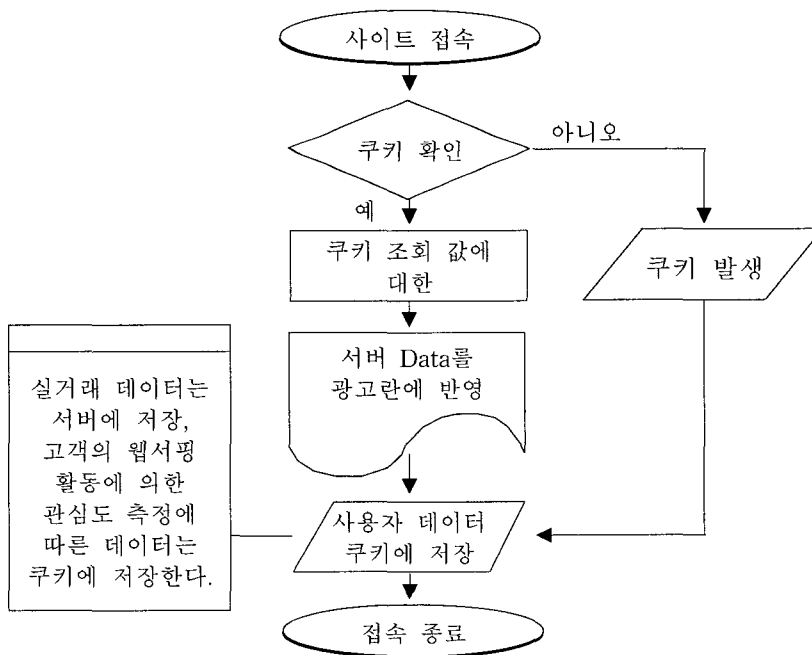
위에서 언급한 사항을 기준으로 본 연구에서 설계된 웹 사이트에서 광고들은 로그인을 거치지 않고 사용자들에 대한 기존의 검색결과를 개인화하여 보여줄 것이다. 즉, 로그인을 통하지 않고 쿠키가 인식하고 있는 IP주소를 기본키로 하여 사용자들의 성향에 알맞는 광고를 연출할 것이다. 본 연구에서 개발된 시스템은 실거래 데이터는 서버에 저장하고, 고객의 웹서핑활동에 따른 관심도(성향)와 관련된 데이터는 IP주소를 식별자로 하여 저장된다.

따라서 사용자가 사이트에 접속하면 기존의 쿠키를 확인하여 쿠키가 존재하면 쿠키와 관련된 서버의 데이터를 호출하여 광고란에 제시하게 되고 결과물을 다시 쿠키의 사용자데이터에 저장함으로써 추후 개인화와 관련된 정보로 사용하게 한



→ : 로그인 중의 개인화 데이터 처리경로
 → : 로그아웃 중의 개인화 데이터 처리경로

<그림 2> 데이터마이닝 에이전트의 데이터 이동경로

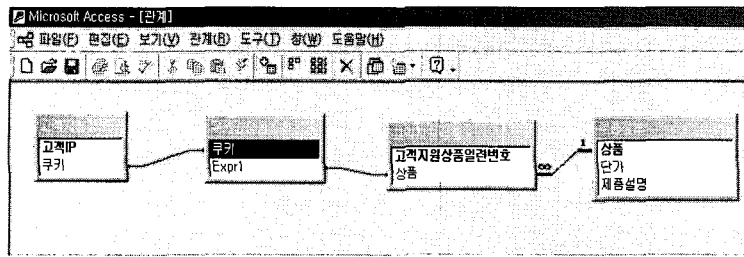


<그림 3> 에이전트 프로그래밍에 대한 순서도

다. 특히, 특정로그인을 거치지 않더라도 사용자들은 기존에 검색된 결과를 이용하여 원하는 검색을 할 수 있으며 사이트에 입장과 동시에 기존 검색 결과는 웹DB에 전송되어 고객의 관심에 따른 광고를 보여줄 수 있도록 설계되었다. 즉 사용자의 웹 활동이 객체화되어 연동하는 데이터마이닝에

에이전트로 인해 IP별로 분류되어 DB서버에 다시 저장된다.

일반적으로 데이터마이닝은 회원가입을 통한 회원ID로 기본키를 갖추고 그에 따른 부수적 데이터를 수집하지만 로그아웃 상태의 데이터까지 수집하기 위해 쿠키가 인식할 수 있는 식별자인 IP주



<그림 4> 배너 광고를 호출하는 DB 관계도

소를 기본키로 참조한다. 그로 인해 사용자 개개인에 대한 차별화된 인터페이스 제공을 가능하게 한다.

배너광고를 호출하는 DB관계도를 <그림 4>에 나타내고 있으며, 배너광고를 호출하는 관계도를 보여주고 있다. <그림 4>는 쿠키를 이용하여 사용자들이 선호하는 상품들을 호출하는 과정을 일례로 보여주고 있으며, 기타 다른 경우에도 이와 같은 정보호출과정이 적용될 수 있을 것이다.

CBMG(customer behavior model graph)는 고객이 사이트내의 링크를 통해서 이동할 수 있는 경로를 보여주며, 보통 사이트맵 또는 카테고리로 나타낼 수 있다. 기업은 고객의 제품구매를 유도하기 위해서 다양한 형태의 콘텐츠를 제공하게 되고, 고객은 특정제품을 구매하기 까지 웹사이트에서 다양한 형태의 행동을 보이게 된다. 따라서 기업은 고객이 제품을 구매하기까지의 일련의 행동양식을 파악할 수 있고, 이러한 정보를 1:1마케팅등에 활용할 수 있다. 본 연구에서 개발된 에이전트 기반의 시스템도 로그데이터를 통해 이와 같은 CBMG정보를 추출할 수 있도록 구축되었다. 특히 CBMG는 접속한 고객에 따라서 여러 가지 행태를 가질 수 있다. 전자상거래 사이트에서 CBMG분석을 통한 로그분석은 고객의 구매행태 등을 파악하는데 높은 신뢰성을 제공한다.

서버에 기록되는 로그파일은 크게 get방식과 post방식의 두가지로 나눌 수 있다. get방식은 url(cs-uri-stem)의 뒤에 질의어(cs-uriquery)를 붙여서 보내는 방식이고, post방식은 질의어를 http헤더안에 포함시켜 전송하는 방식이다. 두가지 방식의 가장 큰 차이점은 질의어를 보내는 방식의 차이인데, get방식으로 질의어를 보낼 경우에는 로그파일에 cs-웹로그 데이터분석에 필요한 데이터 항목은 로그파일 형식에 정의된 기본값을 사용한

다. 정의된 기본값중 필요한 항목은 날짜, 시간, client ip주소, uri-stem, uri-query이다. uri-stem이란 client가 서버에 요청한 파일명을 의미하고, uri-query는 client가 서버에 요청한 파일에 따른 질의어를 의미한다. 본 연구에서는 웹로그분석의 효율성을 위해서 전체프로세스의 각 기능을 get방식으로 정의하였다. 이와 함께 레거시시스템 데이터는 프로세스의 기능에 따른 데이터를 사용하게 된다.

초기의 웹로그분석은 주로 고객이 웹사이트에서 보인 행위를 분석하고, 이를 웹사이트의 디자인전략등으로 이용하였다. 선행연구들[4,5]은 로그분석을 통해서 사용자의 패턴을 분석하여 사용자특성에 중점을 둔 차별화된 웹페이지 디자인전략을 제안하였다. 이를 위해서는 웹사용자의 패턴을 분석해야 하며 사용자의 접속패턴을 클러스터링하고 일정한 규칙을 제공하는 방법론을 개발해야 한다. 또한 확률기반의 군집 프레임워크를 제안하고 고객들의 웹사이트 방문형태를 다양하게 분류하여야 한다.

온라인거래가 활성화되면서 웹로그분석은 고객의 구매행위에 따른 마케팅수단으로 이용될 수 있다. 일대일 마케팅에서 로그분석에 필요한 사전처리과정을 설계하여 이를 통해 의미있는 규칙을 생성할 수 있다. 또한 개인의 이력경로와 군집의 이동경로를 관찰, 예측함으로써 고객군집에 대한 적절한 마케팅 프로모션을 수립할 수 있다.

인터넷 경제의 급격한 변화속에서 고객은 더욱 높은 수준의 서비스를 기대하게 되고, 요구와 취향 역시 다양해지고 있다. 특히 eCRM은 기존의 오프라인 환경의 인터넷 채널로 발전된 것으로 적극적 마케팅활동의 도구로서 인식되기 시작했다. eCRM은 인터넷을 활용한 온라인상에서 고객관리를 통해 고객의 충성도를 높이고, 고객으로부터의 수익

을 최대화하려는 일련의 활동이라 할 수 있다. e-CRM에서 중요한 실행적인 활동은 고객만족을 통한 고객가치의 향상을 목적으로 한다. 따라서 고객만족은 마케팅개념의 핵심으로서 기업이익의 창출과 직접적인 관련을 가진다. 이와 같은 의미에서 본 연구에서 개발된 웹로그분석과 관련된 tool들을 적용하면 고객충성도 향상에 효율적인 방법이 될 것으로 기대된다.

4. 결론

본 연구에서는 데이터마이닝 에이전트를 이용하여 기업이 고객을 관리함에 있어 개별화된 인터페이스를 지원하여 차별성을 둘 수 있는 방법들을 언급하고 있다. 즉 온라인상에 일어나는 행위들을 분석하여 기업의 이윤추구와 방향제시에 활용될 수 있도록 하는 것을 주된 목적으로 하였다. 본 연구에서 개발된 시스템은 현재 국내의 취업서적 전문 쇼핑몰에 일부 이용되고 있으며 거기서 소비자들 이 주로 찾는 서적을 조회하여 웹운영자의 재고 관리에 대한 의사결정을 도와주는 역할을 담당하고 있다.

본 연구에서 제안하는 시스템은 기존의 데이터마이닝 방법에 비해 유연성과 적용성 측면에서 유리하다. 도구기반의 시스템은 서버에 입력될 로그파일을 기다려야 하지만 에이전트기반의 시스템은 세션 내에서 쿠키정보를 추출하고 클라이언트의 데스크 탑 컴퓨터를 서버 대신 사용한다. 물론 클라이언트에서 조회된 정보와 어울리는 값은 웹서버에서 정보를 가져오지만 웹서버의 데이터는 쿼리를 통한 재사용만 일어날 뿐 새로운 데이터가 자주 발생하는 것은 아니다. 그러므로 시스템에 대한 이해가 풍부한 설계자라면 신뢰성이 있으며 고객이 발생시킨 데이터의 손실이나 중복이 없는 에이전트를 손쉽게 설계할 수 있다. 또한 에이전트방식의 데이터마이닝은 본 연구에서 언급되는 고객에 따른 개인화 인터페이스의 구조설계 뿐 아니라 로그분석과정을 색다른 방법으로 모색하여 의사결정시스템 및 전문가시스템의 설계에도 도움이 될 수 있다.

로그파일은 사이트를 방문한 사람들의 방문 데

이터를 서버가 인식하여 차곡차곡 방문자 기록을 정리한 하나의 데이터이다. 이러한 데이터는 우리가 인식할 수 있는 문자형태로 되어있는 것이 아닌 암호처럼 거의 대부분이 숫자로 되어있다. 때문에 잘못하면 로그분석을 시스템적 관점에서 접근하는 경우가 있다. 따라서 분석목적을 명확하게 설정하고 분석결과와 전략적 활용방안을 명확하게 정의하는 것이 필요하다. 본 연구에서 논의된 로그데이터의 주된 전략적 활용 방안은 다음과 같다.

첫째, 고객들의 성향을 파악하여 고객들이 원하는 맞춤 서비스를 제공할 수 있다. 즉 고객의 다양한 요구를 예측하여 새로운 사이트를 개발하거나 새로운 시장 기회를 창출하게 된다.

둘째, 사이트의 정보설계를 경제적으로 할 수 있다. 로그 분석을 통해 잘 찾지 않는 콘텐츠는 서버 화면으로 돌리거나 폐기하는 방법을 통해 무분별하게 콘텐츠를 확장해서 방문객을 혼란스럽게 하는 일을 방지할 수 있다. 특히 사이트 디자인에서는 탐색설계가 중요한데 로그분석을 통해서 사이트 트리모델링에 참고할 수 있다.

셋째, 특정 고객에 대한 집중적인 마케팅 및 광고전략이 가능하다. 마케팅 및 광고전략은 기존의 매스마케팅이나 매스광고처럼 무차별적으로 광고 집행을 하는 것이 아닌 방문자의 방문경로 및 페이지뷰, 임프레션이 많은 페이지에 집중적으로 광고를 수행할 수 있다.

마지막으로, 최적의 환경에서 사용자들이 사이트를 탐색하고 방문하도록 서버 및 회선 등의 기술적 자원을 계획하거나 수립할 수 있다.

일반적으로 eCRM 전략수립은 환경분석, 고객분석, eCRM 전략방향설정, 고객에 대한 오피결정, 개인화 설계 및 커뮤니케이션 설계로 진행된다. 환경분석은 eCRM을 구축하기 전에 eCRM을 통해 얻고자 하는 정확한 목표설정과 시장의 환경에 따른 적절한 고객대응을 위한 시장분석작업이다. 고객분석은 회사가 보유한 고객들이 어떻게 구성되는지를 파악하는 작업이다. 전략방향설정은 기업이 eCRM전개를 통해 궁극적으로 무엇을 얻기 위하여 어떤 활동을 전개하는지에 대한 방향설정이며, 오피결정은 대상고객의 분석을 기반으로 전략방향에 따른 고객관계관리를 위한 구체적인 활동을 수립하는 것이다. 개인화 설계는 고객의 인적특성 및

구매이력에 따른 상품 및 정보를 제공하는 것이다. 마지막으로 커뮤니케이션설계는 실질적인 고객접촉채널을 기반으로 고객을 획득, 개발, 유지할 수 있는 고객 커뮤니케이션을 설계하는 것이다. 따라서 본 연구에서 설계한 에이전트기반의 상품정보 제공은 사용자들의 구매이력을 감안한 개인화설계를 위한 한가지 방법론이 될 것이며, 궁극적으로 eCRM구현의 정당성을 위한 한부분으로 생각될 수 있을 것이다.

참 고 문 헌

[1] Amarasiri, R. and D. Alahakoon, "Building a cluster of intelligent, adaptive web sites," Neural Computing and Applications, Vol.13, No.2, 2004, pp.149-156.

[2] Andrews, J.H. and Y. Zhang, "General test result checking with log file analysis," IEEE Transactions on Software Engineering, Vol.29, No.7, 2003, pp.634-648.

[3] Breeding, M., "Analyzing web server logs to improve a site's usage," Computers in Libraries, Vol.25, No.9, 2005, pp.26-29.

[4] Brown, S.A., Customer relationship management, John Wiley & Sons, 2000.

[5] Forte, D.V., "The "Art" of log correlation: tools and techniques for correlating events and log files," Computer Fraud and Security, Vol.24, No.8, 2004, pp.15-17.

[6] Garatti, S., S.M. Savaresi, S. Bittanti and L.L. Brocca, "On the relationships between user profiles and navigation sessions in virtual communities: A data-mining approach," Intelligent Data Analysis, Vol.8, No.6, 2004, pp.579-600.

[7] Jamali, H.R., D. Nicholas and P.A. Huntington, "The use of users of scholarly e-journals: A review of log analysis studies," Aslib Proceedings, Vol.57, No.6, 2005, pp.554-571.

[8] Manber, U., A. Patel and J. Robinson, "Experience with personalization of Yahoo!," Communications of the ACM, Vol.43, No.8, 2000, pp.35-40.

[9] Riecken, D., "Personalized communication networks," Communications of the ACM, Vol.43, No.8, 2000, pp.41-43.

[10] Schweitzer, D., "Don't ignore lowly log analysis," Computerworld, Vol.39, No.4, 2005, pp.4-5.



손 달 호 (Dal-Ho Son)

- 경북대학교에서 학사
- 미국 Texas Tech에서 석사와 박사학위를 취득
- 현재 계명대학교 경영정보학과에 재직.

• 관심분야: 전자상거래, DSS 등.

홍 덕 훈 (Duck-Hoon Hong)

- 계명대학교에서 학사 및 석사 학위를 취득
- 현재 박사과정에 재학
- 관심분야 : 시스템개발 및 전자상거래 등.