

웹사이트 로그 분석과 사례기반의 전자상거래 에이전트 시스템 (An e-Commerce Agent System based on Web Site Log Analysis and Case Base)

나 윤 지* 원 동 유** 이 순 애*** 조 용 환****
(Yun-Ji Na) (Dong-You Won) (Sun-Ae Lee) (Yong-Hwan Cho)

요 약

전자상거래시스템의 효율적인 운영과 관리를 위해 더욱 많은 노력이 요구되고 있으며 고객의 요구에 대해 가장 적절한 정보를 제공함으로써 만족을 극대화할 수 있어야 한다. 이를 위해 많은 지능형 에이전트 기술을 사용한 전자상거래 시스템이 도입되고 있다. 본 논문에서는 전자상거래시스템에서 CRM 지원을 위한 사례기반 추론과 웹 로그 분석을 기반으로 한 다중 에이전트 시스템을 제안하였다. 본 논문에서 제안한 시스템은 로그분석을 통해 트래픽과 시스템 정보의 분석과 사례기반 추론을 통해 특정 조건에서 고객의 요구에 대해 적절한 정보를 제공할 수 있다.

ABSTRACT

Electronic commerce system must provide information satisfying customers. To do this, many kinds of studies are being advanced actively about electronic commerce support system using intelligent agent technology. This paper suggests electronic commerce system using intelligent multi agent based on Case base and Log analysis. This system can provide suitable information to customers by using knowledge and traffic and system information analysis.

1. 서론

전자상거래는 기업, 조직, 그리고 개인의 제반업무 행위를 위한 모든 유형의 정보를 컴퓨터 네트워크를 통하여 전부 또는 일부를 처리하는 거래 방식이다. 전자상거래의 유형은 거래 주체별 관계를 중심으로 개인과 기업이 전자적 매체를 통신망과 결합하여 재화나 용역을 거래하는 제반 행위인 기업과 소비자간 전자상거래(B to C: Business to Customer

electronic commerce)와 기업 정보시스템을 활용하여 기업 내 데이터를 전자적인 수단과 선진화된 방식으로 업무를 수행하는 체제인 기업내 전자상거래(B to E: Business to Enterprise electronic commerce), 기업이라는 경제 주체들이 동종·이종, 협력·하청 관계로 가상 공간에서 전자적인 수단으로 상호 제품이나 서비스, 또는 정보에 관한 거래 관계를 맺는 기업간 전자상거래(B to B: Business to Business electronic commerce) 등으로 분류할 수 있다[3]. 이

* 정희원: 대전보건대학 컴퓨터정보처리과 초빙교수

논문접수 : 2002. 2. 20.

** 정희원: 충북대학교 전기전자 및 컴퓨터공학부(BK21 Post Doc.)

심사완료 : 2002. 3. 14.

*** 정희원: 세명대학교 강사

**** 정희원: 충북대학교 전기전자 및 컴퓨터공학부 교수

러한 전자상거래는 국내외를 막론하고 이미 활발하게 진행되고 있으며 그 거래 규모 또한 계속 증가하고 있고 초고속 통신망과 가상현실기술과 같은 관련 기술이 보편화되면 전자상거래의 범위 및 규모는 더욱 확장되고 있다. 이러한 규모의 확대와 함께 전자상거래시스템의 효율적인 운영과 관리를 위한 노력이 요구되고 있으며 고객의 요구에 대해 가장 적절한 정보를 제공함으로써 만족을 극대화할 수 있어야 하지만 현재의 많은 전자상거래시스템은 정보를 제공하는 정보 제공사와 고객이 원하는 정보를 만족할 만한 수준으로 제공하고 있지는 못한 실정이다. 효율적인 전자상거래시스템을 위해서는 편리한 인터페이스를 제공해야하며 정보 제공사와 고객을 대상으로 한 고객관계 관리기법(CRM: Customer Relationship Management)의 도입이 필요하다.

로그분석은 웹을 기반으로 한 고객관계관리에서 중요한 기법이며 전자상거래시스템에서는 로그분석을 통해 트래픽(traffic)과 시스템의 효율적인 분석은 가능하나 방문자에 관한 세밀한 정보는 구하기 어려운 단점이 있다. 또한 사례기반 추론의 지식 표현은 고객들의 특별한 상황의 사례들에 대한 추론으로 나타나게되지만 시스템과 트래픽의 분석은 어렵다.

본 논문에서는 전자상거래시스템에서 효율적인 고객관계관리 기법의 도입을 위해 웹 로그 분석을 통한 지능형 사례기반 추론 기반의 다중 에이전트 시스템을 제안하였다. 본 논문에서 제안한 시스템은 사례기반 추론과 로그분석을 이용한 복합적 에이전트 시스템으로 프로토타입을 통해 분석하였다.

2. 관련연구

2.1 CRM

CRM에 대한 정의는 다양하지만 가트너 그룹에 의하면 신규 고객 획득, 기존 고객 유지 및 고객 수익성을 증대시키기 위하여, 지속적인 커뮤니케이션을 통해 고객 행동을 이해하고, 영향을 주기 위한 광범위한 접근이라 할 수 있으며 CRM을 위해서는 고객에 대한 학습단계와 고객에 대한 대응 단계의 반복이 이루어진다. CRM은 다양한 고객 데이터를 내부,

외부 등으로부터 추출하여 업계에 알맞은 논리 데이터 모델을 기반으로 데이터웨어하우스(Data Warehouse)를 구축하고, 데이터웨어하우스를 기반으로 다양한 리포트의 작성과 분석 및 예측을 실시한다. 이를 기반으로 마케팅 기획과 실행을 자동으로 수행할 수 있으며 이와 같은 과정을 통하여 세일즈, 고객 유지 및 다양한 접촉 채널을 통하여 실시하면서 고객의 관계를 관리하며, 최적의 고객에게 최적의 정보를 최적의 타이밍에, 최적의 채널로 판매를 하는 관계를 관리하도록 한다.

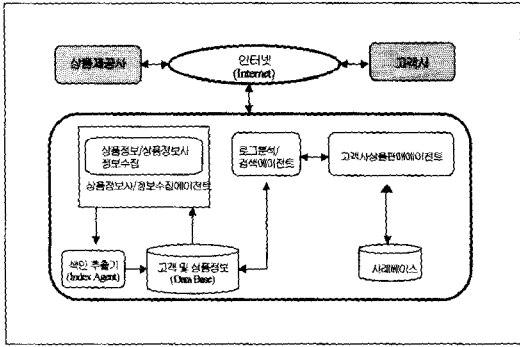
2.2 에이전트

에이전트(Agent)란 사용자를 대신하여 하는 작업을 자동으로 해결해주는 소프트웨어[2,4]라고 할 수 있다. 따라서 전자상거래 에이전트는 전자상거래 상에서의 고객, 판매자, 중간인 등의 전자상거래 참여자를 대신하여 업무를 수행하는 소프트웨어이다. 지능형 에이전트는 사용자가 직접 수행하여야 하는 각종 작업을 대신 수행해 주는 소프트웨어이다. 지능형 에이전트는 사용자가 요구하는 작업을 이해하고 이를 효과적으로 수행하기 위한 계획 기능을 필요로 하며 복잡한 작업을 효율적으로 수행하기 위해서 여러 개의 에이전트가 협동으로 문제를 해결하기 위한 구조가 요구된다. 따라서 지능형 에이전트는 다중 에이전트가 서로 협동하여 가장 효과적으로 사용자의 요구를 충족시킬 수 있는 계획을 수립하고 이를 실현하는 소프트웨어이다[2,4,6]. 전자상거래시스템에서의 에이전트는 각 기능별 특성에 따라 다양한 방법으로 구현 될 수 있으면 개방성 및 상호운용성, 변화에 대한 적절한 대응 등 여러 가지가 고려되어야 한다. 지능형 에이전트는 해당조건에 상태파악을 위한 정보 수집, 정보 처리, 지식 추출, 의사 결정, 의사 결정 수행 및 추론 등을 하게 되며, 그 수행하는 기능에 따라 감시, 학습, 쇼핑, 정보 추출 에이전트로 구분된다. 이러한 전자상거래를 위한 에이전트에는 BargainFinder 등이 있다[7,8,9].

2.3 추론 기법

사례기반추론(Case-Based Reasoning)은 자연적인 학습기능을 갖으며 사례기반 추론에서 지식의 표현

은 특별 상황의 사례들을 사용한 추론으로 나타나게 된다[2]. 사례기반 추론에서 사례(Case)란 어떤 영역의 문제에서 특정한 상황을 표현한 것으로 인공 신경망에서의 형태에 해당된다. 그리고 사례베이스(Case Base)는 특정한 사례들을 모아 데이터베이스화한 것을 규칙기반 전문가 시스템에서 지식베이스에 해당한다. 규칙기반추론(Rule-Based Reasoning)은 전문가의 경험에 의해 발견된 지식을 생성 규칙(Production Rule)의 형태로 표현한 것을 말하며 문제 영역의 규칙들을 전문가로부터 추출한 다음 그것을 정리하여 규칙 베이스를 구현하며 이를 추론하는 것을 말한다. 여기서의 지식 표현은 If-Then규칙으로 표출한다. 사례기반 추론은 규칙기반 추론과 마찬가지로 경험에 의해 스스로 발견한 지식을 이용하므로 내용이 잘 정리되지 않거나, 규칙추출이 어려운 분야의 문제를 해결하는 데 특히 유용하게 사용되고 있다[10].



[그림 1] 시스템 구조

[Fig. 1] System structure

3. 시스템의 설계

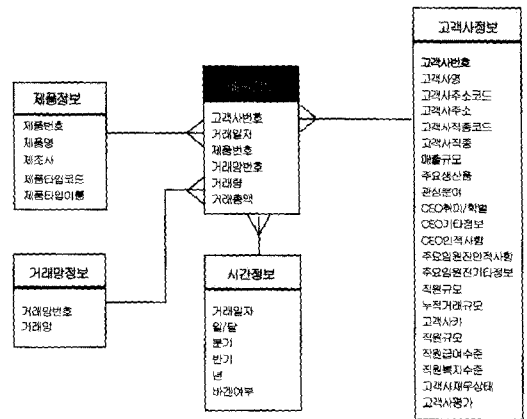
3.1 시스템 구조 설계

시스템의 구성은 [그림 1]과 같다. 정보수집에이전트는 정보나 이를 제공하는 고객 정보를 관리한다. 고객 및 정보 데이터 베이스와 비교하여 저장되어 있지 않을 때에는 해당 정보를 저장하고 색인 추출 에이전트에 의해 색인을 추출하여 키워드 정보를 제

장한다. 색인추출에이전트는 정보로부터 주제별, 품목별, 종류별, 크기별 등으로 색인을 추출한다. 검색 에이전트는 키워드 검색을 지원하는 검색 에이전트이다. 따라서 입력된 정보들은 색인 정보를 이용하여 검색이 가능하며 고객의 요구사항과 특성에 따라 구매할 정보를 검색한다.

3.2 로그분석

인터넷 사용자들이 웹사이트를 방문하고, 제품이나 서비스를 구매하는 과정을 통해 발생하는 데이터를 로그(Log)라고 한다. 이러한 로그는 로그 데이터의 수집 원천에 따라 웹사이트에서 수집되는 데이터와 개별 사용자에게 수집되는 데이터로 구분할 수 있다.



[그림 2] 데이터베이스 설계

[Fig. 2] Database design

첫 번째, 웹서버는 인터넷 사용자가 접하면 자사의 웹사이트에 접속하면 자동적으로 방문횟수, 페이지 뷰 등에 관해 데이터를 저장한다. 이를 통해 방문자의 IP주소를 통해 개별 방문자를 식별하며, 접속 경로를 유추할 수 있으며 방문자가 웹사이트의 내부로 탐색할 경우 페이지별 방문 수 및 페이지에서의 체류시간 등을 알 수 있게 해준다. 이를 통해 웹서버 로그를 분석할 수 있다. 두 번째, 사용자 트래킹 로그이다. 사용자 트래킹 로그는 개별 사용자가 특정 소프트웨어를 설치해야 로그 데이터를 수집할 수

있다. 일반적으로 사용자 트래킹 정보를 수집하기 위해서는 개별 사용자의 승인 하에 웹 서핑활동에 따라 발생하는 로그를 수집하게 된다. 승인과정에서 개별 사용자는 패널이나 고객의 형태로 직종, 주소, 매출규모, 종업원 규모, 주생산품, 소재지성별, 주 관심정보, CEO정보, 주요입원진 정보 등에 관한 정보를 기입하게 된다. 따라서 웹 서버로그와는 달리 사용자의 로그정보와 고객정보를 통합하여 분석할 수 있다. 로그분석은 로그 데이터를 이용하여 트래픽을 파악하고, 이 트래픽이 지닌 의미를 분석해 나가는 과정이라고 할 수 있다. 확장된 의미의 로그 분석은 단지 로그 데이터뿐 아니라, 웹사이트에서 보유하고 있는 고객등록정보, 구매정보, 외부환경정보 등을 복합적으로 사용하는 분석을 말한다. 이러한 분석을 통하여 사용자 특성별로 웹사이트의 이용, 구매에 대한 보다 폭 넓은 분석이 가능해진다. 로그분석의 절차는 다음과 같다.

- 단계1) 로그분석 항목을 선정한다. 로그 분석을 하기 전에 웹사이트를 효율적으로 운영하기 위해 필요한 정보에 대해 정의해야 한다.
- 단계2) 로그 데이터를 수집한다. 필요 정보에 대한 정의를 한 후에, 이에 필요한 분석 항목을 결정하여 로그 데이터를 수집해야 한다. 이는 전처리 과정이나 데이터마이닝을 포함한 통계분석을 할 때 리소스가 감소되며, 로그 데이터의 크기를 줄일 수 있다.
- 단계3) 전처리 과정을 통해 데이터를 선별, 정제 및 변환한다. 로그 데이터가 저장되면, 이를 분석 가능한 데이터로 변환하는 작업이 필요하다. 로그에 저장되는 것은 원시적인 형태의 데이터다. 따라서 단순히 로그 파일만을 가지고 분석할 수 있는 범위는 한정적이다. 또한 데이터 분석에 있어서 가장 중요한 요소 중의 하나는 데이터의 질로서, 초기 로그 데이터는 이러한 요건을 충족시키지 못하고 있는 것이 대부분이다. 전처리

과정은 [그림 3]과 같다.

단계4) 통계 분석 및 웹 데이터마이닝을 통해 로그 분석을 수행한다.

단계 3에서 데이터의 선택(Cleaning Data)은 데이터 현황 조사 분석에 필요하지 않은 아이템의 정제 과정으로, 일반적으로 광고 효과 분석에 필요한 부분을 제외한 gif, jp(e)g, map 등의 로그 파일을 정제하게 된다. 이러한 과정을 통해 로그 데이터의 용량이 축소된다. 일반적으로 로그 분석을 할 경우 개별 사용자의 궤적에 관한 정보를 필요로 한다. 사용자 인식에서는 일반적으로 IP Address와 브라우저의 종류를 이용하여 개별 사용자를 식별하게 된다. 세션 인식(Session Identification)은 사용자의 시간초과 여부를 점검하여 정제하는 과정이다.

일반적으로 로그분석을 할 경우, 웹사이트 내에서 체류 시간이나 방문 경로를 파악할 경우 한 번의 방문으로 발생한 데이터를 분석하게 된다. 또한 유동 IP나 프락시 서버를 사용할 경우 발생하는 문제인 여러 사람이 동일한 IP를 사용하는 것을 구분할 필요가 있다. 따라서 웹사이트를 방문하여 일정시간을 경과한 후에 발생하는 움직임에 대해서는 개별 사용자로 구분할 필요가 있다. 경로 완성(Path Completion)은 로그에 기록되지 않은 궤적 정보를 연결하는 과정이다. 로그에 기록되지 않은 경우는 back 또는 forward 버튼을 눌러 이동한 경우에 발생하게 된다. 포맷 전환(Formatting)에서 데이터 전환은 통계 분석이나 데이터마이닝 분석에 필요한 데이터 포맷으로 전환하는 과정이다.

사전에 정의된 필요 정보와 기법에 적합하도록 데이터의 항목과 기간을 선택해야 하며 불필요한 필드를 제거하고 데이터의 타입 및 길이 등을 변환하며 서로 다른 포맷을 통일시키고, 계산된 값을 추가하여 코드를 변환시킨다.



[그림 3] 전처리

[Fig. 3] Preprocessing

3.3 에이전트의 구조

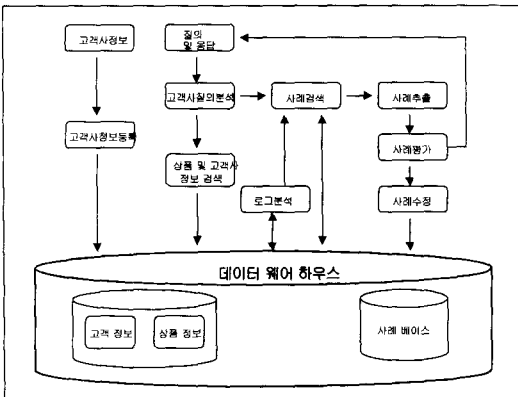
질의 분석은 고객을 식별하고 고객의 질의를 분석하여 질의로부터 키워드를 추출하며 검색 단계는 고객의 정보와 키워드를 이용하여 사례 베이스로부터 관련된 사례를 검색한다.

검색된 사례 중에 고객의 정보와 질의로부터 추출된 정보가 일치하면 해당 사례를 선정하고 일치하는 정보가 없는 경우에는 유사값을 구하여 가장 유사한 사례를 선정한다. 검색을 통해 검색된 사례의 유사값이 너무 낮아 고객이 만족할 만한 사례를 찾아 고객의 질의에 맞는 정보를 검색하는 검색 단계에서 검색엔진을 통하여 고객에게 맞는 정보를 제공하고 유사값이 높거나 유사값이 같은 경우가 발생하지 않는다면 고객의 질의에서 추출한 키워드에 해당하는 정보를 데이터 베이스 시스템에서 검색하여 그 정보에 대한 가장 알맞은 정보를 제공한다. 사례수정은 고객이 최종 구매한 동일한 사례가 사례베이스에 존재할 경우 사례의 회수를 증가시키고, 새로운 사례일 경우 사례베이스에 추가시킨다. 즉 고객이 정보를 구매하지 않았을 경우 사례베이스의 내용은 변화가 없으나 구매하였을 경우에는 사례베이스의 정보를 수정하게 된다. 정보 등록은 새로운 고객일 경우 고객의 정보를 고객 정보 데이터베이스에 추가한다. 에이전트의 구조는 [그림 4]와 같다.

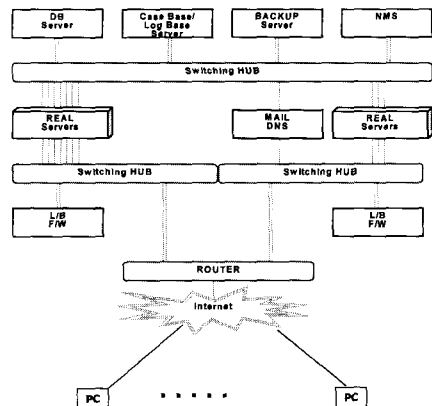
4. 결론

시스템의 프로토타입은 Windows 2000 Server 환경을 이용하며, 웹서버는 IIS5.0을 사용하였고 VBScript와 서버스크립트 언어로는 ASP3을 이용하여 MS_SQL 2000 서버에서 실험하였다.

실험 결과 로그분석을 통해 효율적인 트래픽의 분석은 가능하나, 다음과 같은 요인으로 인해 방문자에 관한 세밀한 정보는 구하기 어렵다. 첫째, 개별 사용자를 확인할 수 있는 IP 정보에 대한 불확실성 문제이다. 클라이언트 IP는 사용자가 인터넷을 시작한 위치 정보에 해당된다. 클라이언트 IP정보를 통해 웹사이트에 접속하는 사용자들의 정보를 파악할 수 있다. 하지만 사용자가 유동 IP를 사용할 경우, 또는 프록시서버를 이용할 경우 등에서는 정확한 클라이언트 IP를 파악할 수 없다. 둘째, 사용자가 웹캐쉬(web cache) 기능을 통한 경우 사용자가 그 페이지를 볼 경우에도 로그에는 그 데이터가 저장되지 않는다. 그리고 뒤로 가기(BACK) 통해 이동할 경우에도 행위들이나 사용자 정보가 로그에 기록되지 않는다. 셋째, 쿠키(cookies)를 이용할 경우 한 명의 사용자가 한 PC를 계속 사용한다는 경우를 가정해야만 그 정확도에 대해서 만족할만한 수준이지만 실제 그렇지 않은 경우가 대부분이다. 따라서 고객이 원하는 적절한 정보를 제공하기 위해서는 고객과 물품 등에 대한 세밀한 정보를 포함한 확장된 개념의 로그 분석이 필요하였다.



[그림 4] 에이전트 구조
[Fig. 4] Agent structure



[그림 5] 시스템 구현 모델
[Fig. 5] System model

전자상거래시스템에서 지능형 에이전트에 대한 연구와 웹 로그를 기반으로 한 CRM을 위한 각종 분석 방법은 다양하다. 로그분석의 경우 효율적인 트래픽의 분석은 가능하나, 방문자에 관한 세밀한 정보는 구하기 어렵다. 또한 대부분의 전자상거래시스템은 카탈로그 형태의 시스템과 요구사항 기반의 키워드 시스템을 지원하고 있다. 하지만 효율적인 전자상거래시스템을 위해서는 고객의 요구를 좀더 충실히 반영한 정보의 정보를 제공할 수 있는 지능형 전자상거래시스템이 필요하다.

본 논문에서는 전자상거래시스템에서 CRM 지원을 위한 사례기반 추론과 웹 로그 분석을 기반으로 한 다중 에이전트 시스템을 제안하였다. 본 논문에서 제안한 시스템은 CRM 지원을 위한 로그분석의 장점인 트래픽과 시스템에 대한 정보와, 사례기반 추론의 장점인 자연스런 학습 기능을 통하여 특정 조건에서 고객에게 도움을 주는데 필요한 지식을 에이전트 스스로 습득할 수 있다. 향후 본 논문에서 제안한 시스템의 적용을 통해 로그분석과 지능형 에이전트 기능을 가진 전자상거래시스템의 효용성에 대한 검증이 필요하며, 이를 통해 발전된 전자상거래시스템의 제안이 필요하다.

※ 참고문헌

[1] Katsumi nihoi의 3인 “Expert Guide for help Desk - An Intelligent Information Retrieval System for WWW Pages”, Proceedings of the Ninth International Workshop on Database and Expert Systems Applications, IEEE Computer Society, pp.937-942, 1998.

[2] 성백균 외 2인 “전자상거래를 위한 사례기반 추론의 판매지원 에이전트,” 한국정보처리학회지 논문지 제 7권 제 5호, pp.1649-1656 , 2000. 5.

[3] 임춘성, “e-Business File,” 영진Biz.com, 2000.

[4] 문기영, “에이전트에 기반한 클라이언트/서버 시스템의 모델 및 응용,” 서울대학교대학원 석사학위논문, 1998. 2.

[5] 임진수, “에이전트 기반의 업무 프로세스 모델링에 관한 연구,” 서울대학교대학원 석사학위논문, 1999. 2.

[6] 백혜정, 박영택의 “적응형 에이전트,” 정보과학회

지, 제 15권 제 3호, pp.5-10, 1997. 5.

[7] 이은석, “멀티에이전트 기술의 실세계 시스템으로의 응용,” 정보과학회지, 제 15권 제 3호, pp.17-28, 1997. 3.

[8] 최중민, “에이전트의 개요와 연구방향,” 정보과학회지, 제 16권 제5호, pp.7-16, 1997. 3.

[9] T. Rolf, Wigand and I. Robert. Benjamin “Electronic Commerce: Effects on Electronic Markets,” JCMC, Vol.1, 1999.

[10] 이용희 “사례기반 추론과 규칙기반 추론을 이용한 자동화기기의 고장진단 시스템,” 금융(통권 542호), pp.67-71, 1999. 5.

나 윤 지	현)대전보건대학 컴퓨터정보처리과 초빙교수 뉴욕공과대학(NYIT)대학원 과정수료(Communication ART 전공) 경북대학교 졸업(이학사)
원 동 유	현)충북대학교 전기전자 및 컴퓨터 공학부 Post Doc. 충북대학교 전자계산학과 전산통계 학 전공(이학박사)
이 순 애	세명대학교 강사 충북대학교컴퓨터공학과(박사수료)
조 용 환	고려대학교대학원 이학박사 현)충북대학교 전기전자및컴퓨터공 학부 교수 현)사단법인 한국콘텐츠학회 회장