

클릭스트림과 서버사이드 이벤트에 의한 확장된 웹 로그 처리시스템

강미정^o 조동섭
이화여자대학교 컴퓨터학과
{woodrock, dscho}@ewha.ac.kr

Extended Web Log Processing System by using Click-Stream and Server Side Events

Mi-Jung Kang^o Dong-Sub Cho
Dept. of Computer Science and Engineering, Ewha Womans University

요 약

인터넷 사용자가 급증하고, 인터넷을 통한 비즈니스에 수의 모델에 대한 관심이 높아지면서 방문자별로 맞춤 정보를 제공하는 퍼스널라이제이션이 인터넷 개발자 및 사용자들의 관심을 모으고 있다. 이러한 퍼스널라이제이션을 위해서 전처리과정인 사용자 프로파일 생성 과정을 확장된 웹 로그 처리 시스템을 통해서 구현해본다. 웹사이트 서버의 확장된 이벤트 처리, 즉 사용자의 행위정보를 로그에 포함시켜 로그정보를 웹 로그 서버에 전송하도록 설계하였다. 그리고 이 웹 로그 서버는 웹사이트로부터 얻은 로그정보를 분류하고 저장하여 관리자가 확장된 웹 로그 정보를 쉽게 분석할 수 있다. 이때 데이터베이스 저장 기술로 OLE DB Provider 상에서 수행되는 ADO 기술을 사용함으로써 확장된 웹 로그 처리 시스템을 설계하였다. 확장된 웹 로그 DB를 패턴분석, 군집분석 등의 마이닝(Mining) 기법을 통하여 맞춤 서비스에 대한 사용자 프로파일을 구축할 수 있다.

1. 서론

인터넷을 사용하다보면 다양한 마케팅 방법으로 사용자들의 관심사에 맞춘 서비스를 제공하여 웹사이트의 재방문 횟수를 늘리고, 판매수익을 높이고 있다.

기존의 대중을 대상으로 한 마케팅이 좀더 세분화된 개인에 초점을 맞춰 이루어지면서 웹사이트의 퍼스널라이제이션(Personalization)이 대두되었다.

웹사이트의 퍼스널라이제이션이 제대로 되지 못하면 오히려 고객들의 불만을 사게 되며, 결국 고객은 그 사이트를 떠나게 된다. 현재의 퍼스널라이제이션 솔루션의 한계는 불충분한 데이터를 기반으로 고객에 대한 부정확한 프로파일링과정을 거쳐 고객의 욕구 충족이 불가능한 서비스를 제공하고 있는 것이다.

프로파일 구축을 위한 고객 데이터는 웹, 이메일, 콜센터 등의 다양한 원천에서 발생된다. 완벽한 프로파일 구축을 위해서는 모든 채널에서 발생하는 데이터를 통합하는 것이 필요하다. 그러므로 기존의 웹 데이터를 이용하는 방법으로는 불충분하므로 웹 데이터를 확장해야 한다. 여기서 웹 데이터를 클릭스트림과 서버사이드 이벤트를 통해서 확장할 것이다.

이러한 확장된 프로파일을 통해서 맞춤 서비스에 대한 신뢰도 및 활용성을 높일 수 있다.

그러므로 본 논문에서는 확장된 웹 로그 처리과정을 통한 충분한 프로파일 구축과 적극적인 맞춤 서비스를 제공하기 위한 전처리 과정을 수행할 수 있는 시스템을 제안하고자한다.

2. 관련연구

2.1 클릭스트림(Click-Stream)

잠재 고객들이 경쟁자의 웹 사이트를 클릭 한 번으로 손쉽게 넘나들기 시작하면서 클릭스트림 분석의 필요성은 대두되었다. 자사의 웹 사이트로부터 고객이 떠나지 않도록 하기 위하여 기업은 고객을 웹 사이트에 묶어두는 요소들을 끊임없이 첨가하여야 한다.

이러한 요소들은 고객의 행동과 선호도, 구매 성향 등에 대한 지식을 기반으로 결정되며 고객의 최근 행위를 반영하는 고객 개개인에 맞춰진 역동적인 콘텐츠 제공으로 이어진다[5].

User, Visit, Page View, Click-Stream, Hit의 의미와 그 처리과정은 다음과 같다[3].

- 웹사이트를 사용하는 고객(Customer)은 unique 'user'가 된다.
- User가 사이트를 탐색하는 때마다 'Visit'가 발생한다.(대부분의 상용 웹 분석 툴에서는 단순히 사용자의 request 시간간격으로만 visit를 구별한다.)
- 각각의 visit는 사용자가 본 일련의 페이지들로 구성된다. 즉 각각의 visit는 시간순으로 정렬된 'page view'로 이루어지며, 이를 'click stream' 이라고 한다.
- 각각의 페이지는 body text, images, video files 등의 많은 구성요소를 가지고 있으며, 각각은 'hit'라는 결과를 내고, 따라서 하나의 page view는 많은 hit로 구성된다.

2.2 웹 로그 데이터 획득을 위한 서버사이드 이벤트

온라인상에서 고객과의 일대일 관계를 수립하기 위하여 고객들이 웹사이트에서나 또는 이메일, 타 채널등을 통하여 전달하는 반응들을 효과적으로 모니터링하고 측정, 평가할 필요가 있다.

고객의 행동을 정의할 수 있는 다양한 변수들을 통하여 고객이 웹사이트에서 제공되는 새로운 콘텐츠나 마케팅 오퍼 등에 어떻게 반응할지를 예측하고 평가할 수 있다. 그러므로 본 논문에서는 전자상거래 서버측의 이벤트 즉, 메일링 리스트 등록, 온라인 의견 등록 등을 통하여 사용자의 반응을 웹 로그 서버에 전송하도록 설계하였다.

2.3 기존 웹 서버 로그 파일

전세계적으로 사용하는 웹 서버는 여러 종류가 있는데, 그 중 Apache 50%이상, IIS(Internet Information Server)는 20~30%정도를 차지하고 있다. 기존 웹 서버로그 정보의 단점은 다음과 같다.

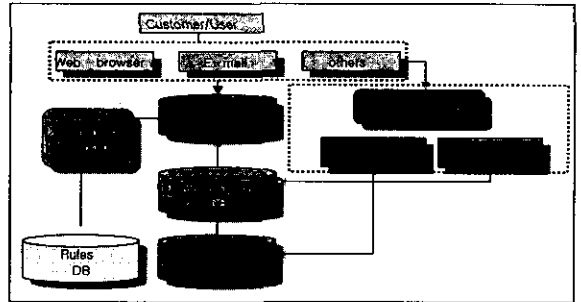
- 로그를 텍스트 파일로 관리하여 검색이 힘들.
 - 로그를 데이터베이스로 저장시 속도가 문제가 됨.
 - 로그에는 출입정보가 주로 기록됨.
 - 로그를 서버 자체에서 처리하므로 서버에 부하가 됨.
- 그러므로 본 논문에서 확장된 로그를 처리하는 서버를 두어 서버 자체에서 처리하지 않게 함으로 서버에 부하를 줄이며 OLE DB Provider상의 ADO기술을 사용하여 데이터베이스를 접근하므로 속도문제를 향상시켰다.

2.4 전처리과정

전처리과정으로 <그림 1>을 보면 웹 사이트를 방문할 때, 처음 등록된 회원정보로 로그인하게 함으로 고객 관련 데이터를 수집하고, 고객과의 접점인 웹사이트, 이메일,

일, 이벤트를 통해서 발생하는 고객 행동 데이터를 통합하여 고객관점의 데이터베이스를 구축한다.

그리고 구축된 데이터베이스를 대상으로 마이닝, 규칙을 적용 분석하여 퍼스널라이제이션 서비스를 할 수 있는 기반을 만드는 과정이다[3].

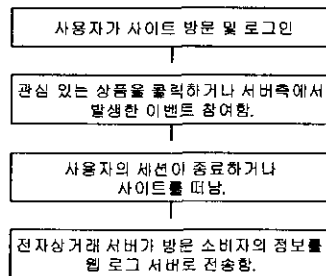


<그림 1> 퍼스널라이제이션을 위한 전처리과정

3. 확장된 웹 로그 분석 시스템의 구현

3.1 전자상거래용 웹 서버의 구현

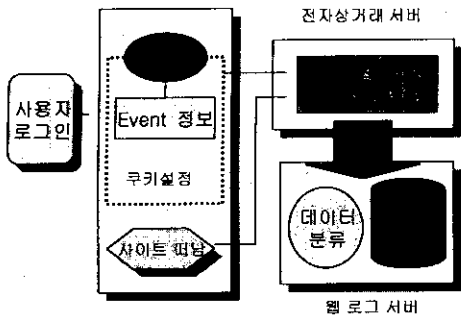
전자상거래 서버는 고객관련 데이터와 고객행동 데이터를 수집할 수 있도록 설계되었다. 웹 상에서의 회원등록 과정을 거친 사용자가 관심 있는 상품을 클릭 하거나 구매하는 경우 쿠키에 정보가 추가되어 사용자가 쇼핑을 마치고 쇼핑몰을 빠져나갈 때, 즉 세션이 종료될 때 저장된 쿠키정보를 웹 로그 분석 시스템으로 전송하는 역할을 한다. 그 수행 과정을 살펴보면 다음과 같다.



고객행동 데이터를 수집하기 위해 관심 있는 상품의 이미지를 클릭한 경우 사용자 이미지 정보와 클릭한 시간이 쿠키에 함께 저장된다. 또 고객이 서버측에서 발생한 이벤트에 참여한 경우 고객의 직접적인 반응을 이벤트 정보와 시간을 함께 쿠키에 저장하여 로그 서버로 전송한다.

즉, 고객 방문 시 IP, 방문시간, 사용 브라우저 정보와 고객의 ID, 이벤트 정보, 클릭한 정보가 함께 저장된다. 다음 <그림 2>는 전자상거래 서버와 확장된 웹 로그 서버사이의 사용자 로그 정보 전송을 나타내는 전체적인

그림이다.



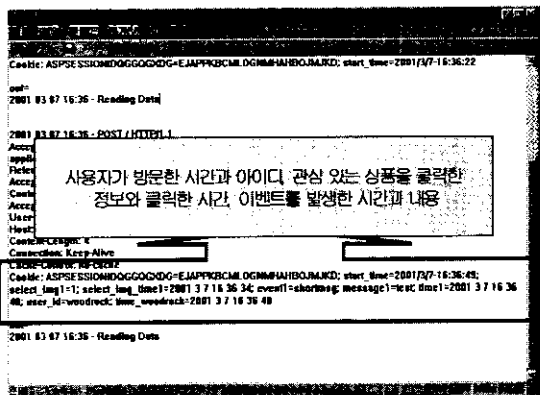
<그림 2> 전체 시스템 흐름도

3.2 확장된 웹 로그 서버의 구현

웹 로그 서버의 작업 순서는 사용자로부터 얻은 확장된 로그 정보를 받고 받은 로그정보를 분류하여 데이터베이스에 저장하게 된다.

데이터베이스에 저장할 때 사용자의 행위 정보와 로그 정보를 OLE DB Provider 상의 ADO 기술을 사용함으로써 속도면이나 기능 면에서 상당히 빠른 액세스로 데이터베이스에 접근하도록 하였다.

다음 <그림 3>는 전자상거래 서버에서 사용자가 관심 있는 상품을 클릭한 정보와 이벤트의 내용이 웹 로그 서버로 전송된 화면이다.



<그림 3> 확장된 웹 로그 정보가 로그 서버에 전송된 화면

3.3 실험 및 평가

현재의 퍼스널라이제이션 솔루션의 한계인 불충분한 데이터를 기반으로 고객의 프로파일을 생성하는 프로파일링과정의 문제점을 들고 이를 해결하기 위한 해결책으로

사용자의 행위정보를 반영하는 이벤트와 클릭정보를 이용하여 사용자의 웹 로그 정보를 확장하고 이 정보를 수집하는 로그 서버를 제안하였다.

확장된 웹 로그 정보를 이용하여 사용자의 욕구 충족이 가능한 서비스를 제공할 수 있는 프로파일 구축을 용이하게 할 수 있다.

4. 결론 및 향후 방향

본 논문에서는 퍼스널라이제이션 솔루션의 전처리 과정인 프로파일 과정을 전자상거래상의 서버사이드 이벤트와 클릭 정보를 이용하여 확장하였다. 확장된 로그 정보를 저장하고 분석하는 웹 서버를 두어 이를 효율적으로 관리하도록 하였다.

로그 파일로부터 얻는 방식은 일반적인 방법으로 안정적으로 데이터의 확보가 가능하나 배치(Batch)로 작업이 이루어져야 하기 때문에 실시간 분석이 불가능하므로 실시간으로 웹 행동 정보를 수집 및 분석할 수 있는 방법을 추가해야 한다. 그리고 보다 효율적인 웹 로그 분석을 통한 사용자 프로파일링을 위해서는 사용자의 등록 정보인 고객 데이터와 웹 로그를 통한 사용자 행위 데이터를 통합한 분석이 필요하다. 또한 웹 로그 분석에 적절한 패턴분석, 군집분석 등 마이닝(Mining) 기법을 적용하여 웹사이트 성격에 맞는 사용자 프로파일을 생성에 대한 연구가 충분히 이루어져야 할 것이다.

5. 참고 문헌

- [1] 박용찬 "e-비즈니스 파워" 2000
- [2] Alex Homer 외 14명 "Activer Server Page 3.0", pp465~515
- [3] HTTP://www.wise.co.kr/
- [4] HTTP://www.microsoft.com/korea/directaccess/
- [5] HTTP://www.oracle.com/
- [6] Alexander Pretschner, Susan Gauch, "Personalization on the Web", 1999
- [7] A.G Buchner, S.S. Anand, M.D. Mulvenna and J.G Hughes, "Discovering Internet Marketing Intelligence through Web Log Mining", 1999
- [8] Wei-Lun Chung, Soe-Tsyrr Yuan, "A Synthesized Learning Approach for Web-Based CRM", 2000, pp43-47
- [9] Robert Cooley, Pang-Ning Tan, Jaideep Srivastava, "Discovery of Interesting Usage Patterns from Web Data", 2000
- [10] Tsvi Kuflik, Peretz Shoval, "User Profile Generation for Intelligent Information Agents-Research in Progress", 1999