

효과적인 추천과 세분화를 위한 트랜잭션 기반 여러 형태 사용자 프로파일의 구축

고 재 진⁺ · 안 형 근^{**}

요 약

쉽게 접할 수 있는 정보의 양이 증가하고 전자상거래가 발전함에 따라, 넓은 정보공간을 축소하기 위하여 추천과 SDI 시스템과 같은 정보 필터링 시스템이 사용되어지게 되었으며, 이에 따라 사용자들은 그들의 요구와 취향에 가장 적합한 정보들을 바로 접근할 수 있게 되었다. 지금까지 다양한 정보 필터링 방법들이 추천시스템을 지원하기 위해 제안되었다. 최근에는 새로운 정보교환 표준으로 떠오르고 있는 XML 문서를 필터링하는 시스템들에 있어서도 다른 접근 방법을 요구하고 있다. 따라서, 본 논문에서 제안하는 시스템은 XML이 가진 구조 정보를 이용하여 여러 형태의 사용자 프로파일을 생성하는 방법을 제안한다. 시스템은 구매와 같은 트랜잭션이 발생하기 전에 사용자 구매 패턴을 분석하기 위해서 필요한 프로파일을 운영자가 직접 정의하는 운영자 프로파일과 이를 적용한 사용자 프로파일의 두 부분으로 구성된다. 운영자 프로파일은 DTD로부터 선택된 항목을 이용하여 DTD를 따르는 문서내의 특정부분을 가리킬 수 있도록 만들어 진다. 제안하는 시스템은 사용자의 구매 행위에 적응력을 가질 수 있도록 보다 정확한 사용자 프로파일을 구축하고, 이와 같은 사용자 프로파일을 기반으로 사용자에게 불필요한 검색과정 없이 필요한 상품 정보를 제공할 수 있도록 한다.

키워드 : 프로파일, XML 필터링, 추천, 세분화, 여러 형태 사용자

The Construction of Multiform User Profiles Based on Transaction for Effective Recommendation and Segmentation

Koh Jae Jin⁺ · An Hyoung Keun^{**}

ABSTRACT

With the development of e Commerce and the proliferation of easily accessible information, information filtering systems such as recommender and SDI systems have become popular to prune large information spaces so that users are directed toward those items that best meet their needs and preferences. Until now, many information filtering methods have been proposed to support filtering systems. XML is emerging as a new standard for information. Recently, filtering systems need new approaches in dealing with XML documents. So, in this paper our system suggests a method to create multiform user profiles with XML's ability to represent structure. This system consists of two parts; one is an administrator profile definition part that an administrator defines to analyze users purchase pattern before a transaction such as purchase happens directly, another is a user profile creation part module which is applied by the defined profile. Administrator profiles are made from DTD information and it is supposed to point the specific part of a document conforming to the DTD. Proposed system builds user's profile more accurately to get adaptability for user's behavior of buying and provide useful product information without inefficient searching based on such user's profile.

Key Words : Profile, XML Filtering, Recommendation, Segmentation, Multiform User

1. 서 론

인터넷과 인트라넷의 발전과 유무선 네트워크의 활성화와 더불어 수많은 형태의 웹 서비스가 있으며, 그 중에서 가장 많은 관심을

가지는 것은 바로 전자상거래(Electronic Commerce) 일 것이다. 전자상거래의 발전으로 정보(거래, 상품 등)의 수는 기하급수적으로 증가하고 있으며, 이러한 정보의 양이 증가할수록 사용자는 각자 자기에게 필요한 정보를 찾기 위해 더 많은 시간과 노력을 투자해야 한다. 이에 따라 방대한 정보 중에서 최적의 정보를 찾고자 하는 사용자의 요구가 증대되고 있으며 이러한 요구를 충족시키기 위해서는 필요한 정보만을 수집하여 사용자에게 제공하는 것이 필요하게 되었다.

* 본 연구는 2003학년도 울산대학교 학술연구 지원으로 수행되었음.

⁺ 정 회 원 : 울산대학교 컴퓨터정보통신공학부 교수

^{**} 정 회 원 : (주)CFIC 기업부설연구소 연구소장

논문접수 : 2006년 5월 3일, 심사완료 : 2006년 9월 1일

수많은 정보 중에서 필요한 정보만을 수집하기 위해서 필요한 것이 정보필터링이다. 최근 전자상거래에서는 정보필터링을 통한 고객의 필요와 요구에 맞는 인터넷 서비스인 개인화 시스템(Personalization System)이 국내외적으로 자리를 잡고 있으며[2, 13], 또한 데이터마이닝(Data Mining)이나 eCRM(electronic Customer Relationship Management) 시스템 등을 이용하여 고객을 유치, 관리 및 서비스를 하고 있다 [1, 4, 6]. 정보필터링에 대한 또 다른 연구로는 수집된 정보를 다시 사용자에게 재구성하여 제공하는 선택적 정보 배포(SDI: Selective Dissemination of Information)를 들 수가 있다 [8-10]. 현재까지 선택적 정보 배포 서비스가 발전하고 서비스 이용자와 대상 데이터도 함께 증가하면서, 다양한 정보로부터 복잡한 질의가 신속히 시행되고 또한 검색된 다량의 정보가 다수의 사용자들에게 빠르게 전달될 수 있도록 하기 위하여 정보검색이나 결과제공 성능의 지속적인 개선이 선택적 정보 배포 시스템의 주요한 요구사항이 되어 왔다. 특히 여러 이용자들이 신청한 다수의 질의를 대용량 데이터베이스에서 검색하는데 많은 시간이 소요되면서 이 시간을 획기적으로 단축할 수 있는 시스템 설계와 개발이 절대적으로 필요하게 되었다.

개인화 및 선택적 정보 배포와 관련된 시스템은 정보를 요구하는 다수의 사용자들에게 합당한 정보들을 시기적절하게 제공하는 것을 목적으로 하고 있다. 이러한 시스템들의 목적을 만족하기 위하여 인터넷을 통한 정보교환의 표준 데이터 형식인 XML(eXtensible Markup Language)이 자리를 잡아가고 있다. 최근까지 증가하는 문서들은 표준화를 위한 작업으로 XML기반의 문서가 주를 이루고 있다. XML은 문서를 사용하는 이유는 문서의 구조정보가 XML 문서 내에 서술되어 있어서 문서를 필터링하는데 그런 구조정보를 활용할 수 있다는 점이다. 또한 XML의 구조정보가 문서의 경로정보이므로 사용자가 원하는 정보만을 쉽게 추출할 수가 있으며, 여러 개의 XML 문서 내용을 모아서 전달하는 일도 가능하다. 또 하나의 장점은 데이터를 표현하는 부분과 프레젠테이션 부분이 분리되어 있다는 점이다. 따라서 하나의 문서가 여러 가지 다른 형태로 제공될 수 있다. 이때 필터링의 단위가 문서가 아니라 문서 내의 일부분을 추출하는 것이라면 XML 문서를 다른 형태의 포맷으로 변환하는 과정에서도 문서의 구조 정보가 필요하게 된다[8, 9].

위 모든 내용으로 볼 때에 전자상거래에서는 사용자의 성향을 정확하게 분석하기 위한 전략시스템을 이용하고 있으나 시스템을 개발하고 관리하는데 많은 비용과 시간적인 소요가 있으며 운영자가 그 시스템을 이해하고 관리하는데 어려운 문제점으로 되고 있다. 또 다른 문제점으로 여러 사용자들에 의한 질의를 대용량 데이터베이스에서 정보를 필터링하여 제공하는데 시간적인 요구사항이 있다. 이런 문제점에서 볼 때에 사용자 요구 정보를 어떻게 잘 분석하느냐가 정말 중요하다 할 수가 있다. 따라서 사용자 요구에 따른 모든 XML 문서의 추출 및 배포를 통하여 적절한 정보를 제공해주는 시스템들은 사용자 요구들의 집합이라고 할 수 있는 사용자 프로파일

(User Profile)에 기반하여 해당 정보가 사용자에게 적합한지를 판단하는 것이 효율적일 것이다. 사용자가 원하는 정보를 사용자 프로파일로 설정하여 서버에 알려주면, 서버는 그 사용자 프로파일을 바탕으로 사용자가 원하는 정보를 제공하거나 배포하게 되는 것이다. 따라서 사용자 프로파일을 하나의 질의 언어라고 할 수도 있을 것이다. 최근 연구까지를 보면 사용자 프로파일을 효율적으로 어떻게 생성을 할 것인가가 대부분이었으며[6, 7, 12, 13, 14, 17], 활용으로는 프로파일을 이용한 전자상거래에서의 추천, 개인화 등과 같은 연구가 있었다[2, 3, 15]. 하지만 이 모든 프로파일은 하나의 규칙에 의한 정적인 면이 강하며 또한 모든 사용자가 그 사이트 내에 만들어진 일관된 규칙에 맞는 서비스만을 추천 받았다.

따라서, 본 논문의 시스템은 사용자 정보 중심의 고정된 하나의 프로파일이 아닌 여러 형태의 프로파일을 생성하는 방법과 그에 따른 활용을 전자상거래에 적용하여 제안하고자 한다. 첫째로, XML이 가지는 스키마 표현 능력과 구조 정보를 이용하여 프로파일을 기술하는 방법을 제안한다. 프로파일은 운영자 프로파일(Administrator Profile)과 사용자 프로파일(User Profile)을 생성할 것이며, 프로파일 정보를 표현하기 위해 XPath(XML Path Language)의 형태로 표현된다. 해당 프로파일은 DTD(Document Type Definitions)로부터 얻어진 문서의 구조에서 운영자가 원하는 부분을 추출하여 프로파일을 생성하고, 프로파일은 트랜잭션(거래) 데이터를 필터링하여 사용자 프로파일을 재구성한다. 둘째로, 운영자 프로파일은 사용자들의 성향 분석을 위한 목적으로 프로파일을 생성하였기 때문에 수집된 정보는 전자상거래 운영에서 빠르게 성향을 분석하기 위한 효과적인 세분화(segmentation) 방법에 적용을 하여 그에 따른 분석 통계자료 및 사용자 그룹에 따른 전체적인 성향 분석이 가능하다. 사용자 프로파일은 전자상거래 이용함에 있어 개인에 맞는 상품을 추천할 때에 활용이 되며, 또한 운영자 프로파일에 의해 파생된 여러 형태의 사용자 프로파일에 의해 상품 추천도 다양하게 이루어 질 수가 있다.

이후 본 논문은 다음과 같이 전개된다. 1장 서론에 이어 2장에서는 본 논문과 관련된 관련연구로 사용자 성향 분석을 위한 전략인 CRM과 eCRM, 이를 위한 세분화와 사용자 프로파일을 이용한 개인화 및 선택적 정보 배포에 대해서 살펴보고, 기존의 사용자 프로파일 생성방법과 문제점 등을 기술한다. 3장에서는 본 논문이 제시하는 여러 형태의 사용자 프로파일의 생성 방법 및 프로파일 생성을 위한 운영자 관리 틀에 대하여 소개를 한다. 4장에서는 여러 형태의 사용자 프로파일을 이용한 시스템의 구현 환경과 방법을 소개하고, 그리고 적용된 결과를 기존 시스템과의 상품추천 비교 및 성능평가로 결과를 분석한다. 마지막 5장에서 논문의 결론으로 끝을 맺는다.

2. 관련 연구

2.1 CRM and eCRM

침예한 비즈니스 경쟁 환경 속에서 성공할 수 있는 비법은 무엇일까? 기업이라면 안고 있는 공통된 관심사이다. 여러 가

지 방법이 있겠지만 궁극적으로는 사용자(소비자) 만족도를 얼마나 증진시키는지가 핵심이다. 이러한 이유에서 CRM에 대한 관심이 급속도로 확대되고 있다.

CRM이란 사용자들에게 더 좋은 서비스를 제공하고 사용자들과의 상호작용을 효율적으로 관리함으로써 기업이 그들의 가장 바람직한 사용자들에게 좀 더 효과적인 차별화(개별적 서비스)를 제시하고, 자원을 효과적으로 배치하도록 하여 기업의 수익을 증대시키기 위한 전략이다. 즉, “The right offer, To the right person, At the right time, Through the right channel”이라 정의할 수 있다[1]. 이 과정에서 온라인 인터넷 환경이 모든 생활을 바꿔 놓기 시작했다. 인터넷이 대두됨에 따라 기업에서는 굳이 방문하지 않고도 고객정보를 인터넷을 통해 손쉽게 획득하게 됐기 때문이다. 여기서 나온 개념이 바로 eCRM이다.

eCRM이란 인터넷에서 상품을 구매하는 사용자나 서비스를 이용하는 사용자를 대상으로 인터넷 기반 기술을 이용하여 신규 사용자의 획득, 기존 사용자의 유지 등을 통하여 기업의 수익성을 최적화 하고자 하는 전략이다. 이미 여러 분야에서 사용자의 이탈 방지와 수익 증대를 위해 CRM과 eCRM을 기업 운영에 적극 적용하고 있다.

2.2 개인화

eCRM에서 뜨겁게 논의되는 이슈인 개인화 전략은 각 사용자별로 맞춤 서비스를 제공하기 위한 핵심 전략중의 하나이다. 콘텐츠를 제공하는 사이트에서 각 사용자의 개인화를 가능하게 하기 위해서 사용자 프로파일을 이용한다. 이러한 각 사용자마다의 프로파일을 이용함으로써 각 사용자들에게 차별화된 서비스를 제공해 사용자들과의 친밀한 관계를 유지하여 충성도를 높일 수 있다. 또한 콘텐츠 사이트에서 개인화 전략은 사용자들의 콘텐츠를 검색하는 시간과 경비를 절약하고 사이트에 대한 만족도를 높여 매출을 극대화하기 위한 전략으로 사용할 수 있다. 개인화 방법으로 크게 규칙기반 필터링, 협업 필터링, 내용기반 필터링으로 구분된다[2].

2.3 선택적 정보 배포

선택적 정보 배포의 의미를 갖는 SDI 서비스는 사용자가 탐색 전략을 입력하여 서비스를 신청하면 데이터베이스가 갱신될 때마다 또는 사용자가 지정한 간격으로 자동으로 질의되어 그 결과를 사용자에게 제공하는 서비스로 정의되는데, 이때 대용량의 데이터베이스로부터 복잡한 데이터 질의과정을 거친 데이터가 다수의 사용자들에게 보다 빠르게 전달할 수 있도록 하는 것이 선택적 정보 배포 서비스의 주요한 요구사항이 되고 있다[8]. 선택적 정보 배포 서비스는 과거 전통적으로 운영자가 수작업으로 개개인의 프로파일에 따라 오프라인 인쇄물을 서비스하는 방식에서부터 출발하여 최근에는 서비스의 신청에서부터 검색 기능, 이용자별 신청정보를 맞춤형 이메일로 제공하는 기능까지 모든 과정이 자동화 기능으로 발전하여 왔다. 즉, 사용자가 웹사이트에서 필요한 정보주제를 입력하여 서비스를 신청하면 신청정보는 데이터베이스를

검색할 수 있는 프로파일로 변환되어 저장되고, 데이터베이스가 갱신될 때마다 또는 사용자가 지정한 간격으로 프로파일을 참고하여 해당 주제를 위한 질의가 대상 데이터베이스에 실행되며, 그 결과는 사용자에게 이메일 등 다양한 매체를 통해 제공되는 방식으로 서비스 일련의 과정이 모두 자동화 된 것이다.

2.4 세분화

데이터마이닝을 이용하여 만들어진 예측모델은 새로운 데이터에 대한 예측을 위해 사용된다. 데이터마이닝 모델을 이용하는 절차와 모델을 만드는 절차는 별개이다. 일반적으로 하나의 모델은 서로 다른 데이터베이스를 평가하기 위해 여러 차례 사용된다. 이러한 데이터마이닝이나 고객세분화를 위해서는 먼저 고객세분화 기준을 가지고 <표 1> 같은 방법으로 고객세분화 변수를 만들어 데이터베이스화하는 것이 중요하다[4, 7].

<표 1> 고객세분화를 위한 변수와 기준

고객 세분화 변수	고객 세분화 기준
고객 통계적 변수	나이, 성별, 직업, 지역, 원급 등
고객 구매 변수	구매 상품, 상품구매액 등
고객 충성도 변수	결제방법, 결제금액 등

2.5 기존 사용자 프로파일 생성 방법

2.5.1 로그인 기반 방법(Login-based method)

시스템은 대부분의 경우 접속하는 모든 상태를 로그파일에 저장한다. 로그 분석이란 이러한 로그파일로부터 시스템에 접속하는 사용자의 호스트 주소, 접속시간, 접속페이지, 에러상태 등의 정보를 분석하여 시스템 상에서 누가, 언제, 무엇을 이용하였는지 한 눈에 알아볼 수 있도록 분석하고 사용자 프로파일을 생성하는 것을 말한다[6]. 로그인 방법의 문제점으로는 로그파일의 용량이 커짐에 따라 시스템의 과다한 리소스 증가, 기간대별 로그파일의 관리 및 그 분석결과에 대한 통합분석이 어렵고, 로그분석 작업 및 로그분석 솔루션 운영을 위한 인력자원의 소요가 필요하다.

2.5.2 인구통계학적 기반 방법

(Demographic-based method)

사용자의 성별, 나이, 직업 등과 같은 인구통계학적 요소에 의해 사용자 유형별 특징을 분석하여 사용자 프로파일을 생성하고 이에 따라 상품을 추천하는 방법이다. 사용자 피드백 정보가 없어도 추천이 가능하여 시스템 초기 구축 단계나 처음으로 시스템에 접속하는 사용자에게도 적용할 수 있다[7]. 하지만 각 사용자에 대한 개별적인 추천이 아니라 사용자 군집으로 구분되기 때문에 개인화된 추천은 효과적이지 못하다.

2.5.3 내용 기반 방법(Content-based method)

내용기반 방법은 사용자의 요구나 사용자로부터 입력된 모

른 정보와 상품에 포함된 내용 정보를 이용하여 필터링한 다음 사용자 프로파일을 생성하는 것이다[12]. 사용자 프로파일과 비교하여 높은 점수를 얻은 아이템을 추천하므로, 사용자가 이미 평가한 아이템과 유사한 아이템만 제공하여 과도하게 집중화 되는 경향을 보인다[17]. 사용자가 개인적인 정보를 입력하지 않으면 개별적인 프로파일을 생성할 수 없다.

2.5.4 협업필터링 기반 방법

(Collaborative Filtering-based method)

협업필터링 방법은 사용자의 상품에 대한 평가와 유사한 선호도를 가진 다른 사용자의 평가에 근거하여 다른 사용자가 선호하는 상품을 추천하는 방법이다. 이 방법은 다양한 사용자 프로파일을 생성할 수 있고 또한 유사한 사용자 간의 프로파일을 찾을 수 있다는 장점을 가지고 있다. 또한 숨어 있는 선호 패턴을 발견할 수 있으면 추천의 정확도가 높은 특징으로 잘 알려져 있다[13, 14]. 하지만 시스템 구축 초기에 사용자로부터 충분한 평가정보가 없을 경우 정확성이 떨어진다. 그리고 상품에 대해서 충분한 평가 결과를 구성하기 어려운 관계로 사용자와 평가정보로 구성된 행렬이 희소성을 가지는 문제가 있으며, 또한 연관관계를 찾기 어렵다.

2.5.5 학습에이전트 기반 방법

(Learning Agent-based Method)

카네기 멜론 대학(CMU)에서 구축한 웹 에이전트인 Personal Webwatcher는 자동으로 모니터링된 사용자 브라우징 행위 정보를 기반으로 관심 정보를 추측한다. 이는 비 감독 학습(Unsupervised Learning)방식의 에이전트로 사용자 관심정보의 진처리 작업과 학습 작업에는 추출된 키워드를 벡터 테이블로 생성 이를 기반으로 TFIDF(Term Frequency Inverted Document Frequency) 알고리즘 및 베이저안 확률(Bayesian Probability)을 적용하여 사용자 프로파일이 구축된다[18]. 이는 사용자 관심 정보를 시스템이 자동으로 추측하기 때문에 사용자에게 필터링을 준다는 이점이 있지만 학습에 입력되는 관심 정보의 신뢰성이 낮다는 단점이 있다.

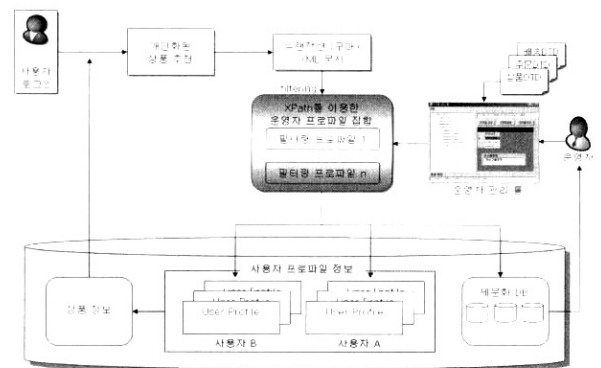
앤디슨 컨설팅 랩에서 개발한 웹 에이전트인 InfoFinder는 사용자의 브라우징 과정과 이어 대한 관심도를 기반으로 학습이 이루어진다는 점에서는 Personal Webwatcher와 동일한 구조를 갖는다[19]. 그러나 사용자의 관심을 학습하는 방식에서 큰 차이점을 갖고 있다. 즉, Personal Webwatcher는 사용자의 관심도를 반복횟수에 따라 일방적으로 결정한다. 이에 반해, InfoFinder는 사용자가 직접 자신의 관심도를 표현할 수 있는 감독 학습(Supervised Learning)을 사용하고 있다[20]. 사용자는 InfoFinder 자체 브라우저를 사용하게 되고 이를 이용해서 사용자 자신의 관심정보를 직접 에이전트 시스템에 표시하게 된다. InfoFinder는 관심 영역내의 중요 정보에 대한 키워드를 단순한 빈번도 위주로 추출하였다.

3. 운영자 관리 툴과 프로파일 생성 관리

본 논문에서는 사용자 분석을 위해 운영자 관리 툴을 이용하여 운영자 프로파일을 생성하게 되며, 이 프로파일은 구매에 의한 XML 기반의 사용자 트랜잭션 데이터를 필터링을 한다. 이러한 필터링을 통하여 사용자 프로파일은 여러 형태로 재 생성되고, 그 결과로 상품 추천과 세분화된 분석결과를 제공하는 방안을 제시한다. 따라서 본 장에서는 운영자 관리 툴에 의한 운영자 프로파일 생성 및 사용자 프로파일 생성에 대해서 알아보고, 이를 이용한 시스템의 전체적인 구조를 기술한다.

3.1 시스템 구조

(그림 1)은 사용자 프로파일을 생성하고 그 생성된 프로파일들이 이용되는 시스템의 전체적인 구성도로서 운영자 부분과 사용자 부분으로 나누어 기술할 수 있다. 운영자 부분에서는 사용자의 유지 및 신규 사용자 유치를 위한 목적으로 사용자들의 구매에 따른 성향을 파악하는 것이 중요하기 때문에 사용자의 정보를 다양하게 필터링하기 위해서 운영자 관리 툴을 이용하여 운영자 프로파일을 생성하여 저장을 한다. 사용자 부분은 전자상거래에서 구매에 의해 발생된 트랜잭션 데이터가 운영자 프로파일의 의해서 필터링 된과 동시에 사용자 프로파일을 재구성하게 되며, 분석을 위한 세분화 데이터 베이스로도 구성이 된다. 또한 사용자 로그인시 재구성된 사용자 프로파일의 의한 개인화된 상품 추천이 이루어진다. 반면에 트랜잭션 데이터가 없는 로그인 사용자는 기본 정보에 의한 상품 추천이 이루어진다. 사용자의 구매 성향은 항상 변화하며 그에 따른 트랜잭션 데이터 또한 변화하기 때문에 본 논문에서 제안하는 사용자 프로파일은 변하는 사용자의 구매 성향에 맞게 제공될 수가 있다.



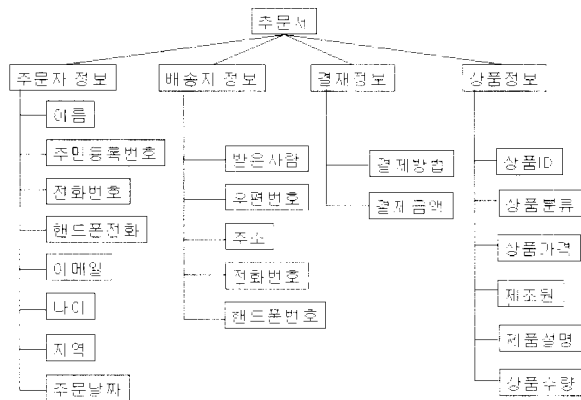
(그림 1) 전체적인 시스템 구성도

3.2 트랜잭션 데이터

사용자 프로파일 생성을 위한 XML 기반의 트랜잭션 데이

터에 대하여 좀 더 자세히 살펴보면, 전자상거래 시스템에서는 고객의 이름, 주소, 생년월일 등과 같은 데이터를 주어진 정보 혹은 사실이라고 하고, 특정기간 동안의 거래 내역등과 같은 데이터를 새로운 사실 혹은 트랜잭션 데이터라고 말한다[3]. 또한 어떤 트랜잭션 데이터를 이용해서 사용자의 행동(Behavior)에 따른 정보를 필터링하여 수집하고 저장한 후 이를 이용해서 새로운 정보를 생성할 수 있는데 이를 트랜잭션 데이터에 의한 사용자 프로파일이라고 말을 한다. 실례를 들면, “사용자가 한 달간 가장 많이 구매한 품목은 셔츠이다.”와 같은 구매 트랜잭션 데이터는 사용자 프로파일로 다음과 같이 표현할 수가 있다. “사용자의 한 달간 구매 선호 품목은 셔츠.”라고 정의를 할 수가 있다. 이러한 사용자 프로파일을 통해서 각 사용자의 소비 패턴 분석, 개인별 선호도 그리고 다양한 정보를 운영자 입장에서 손쉽게 구할 수가 있다.

효과적인 필터링을 위한 구조적인 XML 기반의 거래문서 구조를 보면 (그림 2)와 같이 나타낼 수가 있으며, 이후 모든 내용에서 해당 XML 문서의 구조를 따른다. (그림 3)은 사용자에 의한 상품 구매 후 발생한 트랜잭션 데이터를 XML 문서로 보이고 있다.



(그림 2) 트랜잭션 데이터 DTD 구조

```
<?xml version="1.0" encoding="euc-kr" ?>
<order_form>
  <client_info>
    <name>홍길동</name>
    <ssn_no>860101-111116</ssn_no>
    <tel>052-234-2222</tel>
    <phone>011-111-1111</phone>
    <email>gdhong@xml.ocm</email>
    <age>21</age>
    <area>울산</area>
    <order_date>2006-01-20</order_date>
  </client_info>
  <delivery_info>
    <receive_name>홍길동</receive_name>
    <zipcode>680-111</zipcode>
    <address>울산광역시</address>
    <tel>052-234-2222</tel>
    <phone>011-111-1111</phone>
  </delivery_info>
  <settlement_info>
    <settlement_method>온라인</settlement_method>
    <settlement_price>500000</settlement_price>
  </settlement_info>
  <product_info>
    <product_id>s101</product_id>
    <product_catalog>핸드폰</product_catalog>
    <product_price>500000</product_price>
    <product_maker>삼성</product_maker>
    <product_description>SKY IM-U100</product_description>
    <product_no>1</product_no>
  </product_info>
</order_form>
```

(그림 3) 트랜잭션 데이터 XML 예

3.3 운영자 프로파일

사용자 추천 시스템이 기업 경쟁력 강화에 있어 중요한 이슈로 대두됨에 따라 eCRM 역시 주목 받고 있다. 정확한 사용자의 선호도를 분석하기 위해서 데이터 마이닝, 패턴 인식, 정보 필터링 등의 다양한 방법들이 적용된다. 특히 데이터 마이닝에서 추천 규칙 추출 작업은 데이터마이닝 도구를 사용하여 마케팅 전문가, 데이터 분석 전문가에 의해서 수행되며, 이러한 데이터마이닝 작업을 수행하기 전에 데이터를 적절한 형태로 가공하는 전처리 작업이 필요하고 데이터마이닝을 통해 발견된 패턴 및 규칙을 다양한 목적으로 운영되는 각 시스템에 적합하고 필요한 규칙으로 생성, 유지하기 위해서는 많은 비용이 든다. 이런 문제를 해결하기 위해서 각 시스템 운영자가 직접 자신의 시스템에 적합한 프로파일을 생성하고 유지함으로써 사용자 선호도와 구매 성향을 분석하고 유사한 성향을 가진 사용자를 그룹별 세분화한다. 다음은 운영자 프로파일 생성 방법을 정의하고 (그림 4)와 같이 나타낼 수가 있다.

$$p_n = \{ p_{n_desc}, \{ select_element \} \} \quad n = 1, 2, \dots, n$$

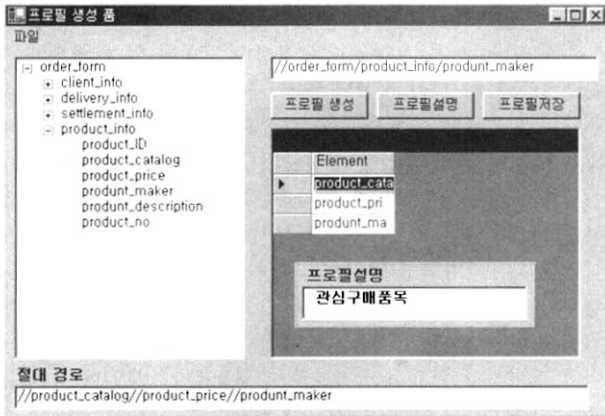
(그림 4) 운영자 프로파일 정의

P_{n_desc} 는 운영자 프로파일 생성 시에 프로파일 간에 구별할 수 있는 프로파일의 설명이며, 메타데이터라고 할 수가 있다. P_{n_desc} 는 운영자 프로파일을 생성 관리하는 운영자 관리 툴에서 입력하여 저장이 된다. $select_element$ 는 XML 기반의 거래문서 항목 중에서 운영자가 성향분석을 필요로 하여 선택한 항목(element)들의 집합이다. 운영자 프로파일 구성을 다음의 예로서 설명하고자 한다. 운영자가 사용자의 성향분석을 위한 지표 자료를 “관심품목” 이라하고, 해당 세부적 분석을 위한 선택 항목으로 (그림 2)의 DTD 구조에서 상품분류(product_catalog), 상품가격(product_price), 제조원(product_maker) 등이라고 할 경우, 운영자 프로파일 구성을 살펴보면 다음과 같다. P_{1_desc} 는 “관심품목”이 되며, $select_element$ 는 {<상품분류>, <상품가격>, <제조원>} 등이 된다. 위 전체적인 운영자 프로파일 구성 표현은 (그림 5)와 같이 표현된다.

$$p_1 = \{ p_{1_관심품목}, \{ product_catalog, product_price, product_maker \} \}$$

(그림 5) 운영자 프로파일 구성 표현

위 운영자 프로파일 생성을 위해서 운영자를 위한 GUI 기반의 프로파일 정의 도구인 운영자 관리 툴을 구현한다. 프로파일 생성을 위한 운영자 관리 툴은 XML 기반 문서를 처리할 수가 있으며, DTD에서 얻어낸 구조 정보를 바탕으로 XML 문서의 각 항목들은 DOM(Document Object Model) 트리 형태로 구성된다. 해당 트리에서 운영자 프로파일 구성에 필요한 항목을 선택할 수 있다. 아래 (그림 6)은 운영자 프로파일 생성을 위한 운영자 관리 툴을 보여주고 있다.



(그림 6) 툴을 통한 운영자 프로파일 생성

운영자는 운영자 관리 툴에서 프로파일 구성에 필요한 항목들을 선택하게 되며, 이때 선택되는 항목들은 문서 내의 경로 정보를 표현하기 위해 XPath 형태로 표현된다. 여기에 제공되는 거래문서는 전자상거래 트랜잭션 데이터 등을 위한 DTD를 따른다고 가정하고 DTD로부터 구조 정보를 추출하여 사용자 프로파일과 대응하게 된다. 또한 운영자 프로파일을 식별하기 위해서 프로파일 생성 과정에서 프로파일 설명을 추가한다. 추가적으로 전자상거래에서 트랜잭션 데이터가 발생하지 않은 기존 사용자나 신규가입 사용자들에 대해서는 구매에 관련된 운영자 프로파일을 생성할 수 없기 때문에 트랜잭션 데이터에 대한 항목별 구분을 위하여 프로파일 적용 범위 영역을 설정을 한다. 트랜잭션 데이터가 발생되지 않은 사용자들은 가입 시에 기본적으로 제공하는 개인정보(성별, 나이, 지역, 직업 등)를 이용하여 프로파일을 생성할 수 있다. 따라서 트랜잭션이 이루어진 고객과 그렇지 않은 고객에 따른 유형별 적용 범위 설정을 보면 <표 2>와 같이 나타낼 수가 있다.

프로파일 적용범위 영역에서 Special 적용 범위는 트랜잭션 데이터가 있는 사용자만 적용이 되며, Public 적용 범위는 트랜잭션 데이터와는 무관한 기존의 사용자들 모두에게 적용될 수 있는 범위를 말한다. 이렇게 생성되어진 모든 운영자 프로파일은 데이터베이스에 저장되고 차후 사용자 프로파일 생성에 영향을 주게 된다.

<표 2> 운영자 프로파일과 적용 범위

Profile ID	Select Element	Scope	Description
P1	<상품분류> <상품가격> <제조사>	Special	관심품목
P2	<나이>	public	나이별
P3	<지역>	public	지역별

Profile_ID	Profile_Desc	scope	element1	element2	element3
P1	관심구매품목	Special	//order_form/pr	//order_form/	//order_form/pr
P2	나이	Public	//order_form/ct		
P3	지역	Public	//order_form/ct		

(그림 7) 데이터베이스에 저장된 운영자 프로파일

3.4 사용자 프로파일

구매가 성사된 이후에 발생하는 트랜잭션 데이터가 운영자 프로파일에 의해 필터링 된 후 사용자 프로파일을 여러 형태로 생성하게 되며, 생성된 각각의 프로파일은 다른 정보를 가지게 된다. 아래의 (그림 8)은 사용자 프로파일 구성 정의를 표현한 것이다.

$$u = \{p_1, p_2, \dots, p_n\} \quad n = 1, 2, \dots, n$$

(그림 8) 사용자 프로파일 구성 정의

이 때 u 는 사용자 ID를 나타내고, P_1, P_2, \dots, P_n 은 사용자에 정의된 사용자 프로파일을 나타낸다. 즉, 사용자는 n 개의 서로 다른 여러 개의 사용자 프로파일을 가질 수가 있다. 사용자 프로파일 생성에 대한 예시로 로그인 아이디가 $u001$ 인 사용자가 전자제품 중에서 TV를 구매하기로 결정한 후 주문서를 작성하고 트랜잭션이 완료되었다고 하자. 이에 따른 운영자 프로파일 생성과 생성된 운영자 프로파일 사용자 프로파일에 어떻게 적용되는지를 본다. 앞 절에서 사용자의 “관심품목”에 대한 프로파일을 생성한다고 할 경우, 운영자는 우선 프로파일 생성 정의 툴을 이용해서 “상품분류”(<product_catalog>)를 선택하고 운영자 프로파일에 대한 설명인 “관심품목”이라는 상세 설명과 Special 영역도 운영자 프로파일로 저장하였다. 위 사용자의 트랜잭션 데이터는 XML 문서로 저장되어 있다. 사용자의 구매로 생성된 XML 문서 형태의 트랜잭션 데이터에는 상품분류 항목인 <product_catalog> 태그를 포함하고 있다. 실제 구매에 의한 완전한 XML 항목 표현은 <product_catalog>전자제품</product_catalog>와 같이 값을 포함하고 있다. 사용자 프로파일을 생성하기 위해서 XML 문서인 트랜잭션 데이터를 필터링한다. 필터링 과정에서는 운영자 프로파일에 정의된 요소 값만 트랜잭션 데이터에서 필터링한다. 그 결과 운영자 프로파일 (그림 5)에 의해 추출되는 값은 product_catalog = “전자제품”이 된다. 생성된 사용자 프로파일의 정의는 $u001 = \{관심품목, <전자제품>\}$ 이 된다. 또한 운영자 프로파일 집합에 정의되어 있는 여러 개의 프로파일을 순차적으로 적용함으로써 여러 형태의 사용자 프로파일을 생성한다. 아래 <표 3>은 각 사용자별로 생성된 프로파일에 대한 정보를 나타내고 있다.

<표 3>에서 보는 것처럼 ID가 $u001, u002$ 인 사용자 프로파일은 트랜잭션 데이터에 의해서 생성된 것이며, ID가 $u003$ 인 사용자는 트랜잭션 데이터가 없으므로 기본적인 개인 정

<표 3> 재구성된 여러 형태의 사용자 프로파일

User ID	User Profile	Scope
$u001$	$P_{관심품목}, P_{나이}, P_{지역}$	Special
$u002$	$P_{관심품목}, P_{나이}, P_{지역}$	Special
$u003$	$P_{나이}, P_{지역}$	Public

보에 의존하는 Public 영역의 범위가 적용된다는 것을 볼 수가 있다.

3.5 세분화 알고리즘

트랜잭션 데이터는 운영자 프로파일에 의해서 필터링되며 사용자 프로파일로 재구성 될 뿐만 아니라 세분화에 대한 데이터베이스도 이루어진다. 운영자 프로파일에서 선택된 엘리먼트는 테이블의 요소로 구성이 되며, 사용자 프로파일 구성 시에 세분화 데이터베이스는 갱신이 된다. 다음 아래 (그림 9)는 세분화 데이터베이스를 구성하기 위한 알고리즘을 나타낸다.

위 알고리즘의 간단한 설명으로 운영자 프로파일에서 선택된 엘리먼트의 이름과 동일한 테이블 이름이 있는지 체크하고, 있다면 엘리먼트의 값이 해당 필드이므로 일정 값을 증가시킨다. 만약에 없을 경우 새로운 선택 엘리먼트의 이름에 대한 테이블을 생성한다. 테이블 생성 후는 위 과정과 동일하게

```

if(select_element_name == DB_table_name) then {
    user id input;
    if(select_element_value == table_field) then
        field_value ++;
    else
        table_field = select_element_value;
}
else {
    creation table_name( = select_element_name)
    user id input;
    table_field = select_element_value;
    field_value ++;
}
    
```

(그림 9) 세분화 데이터베이스 알고리즘

user_id	자켓	셔츠	팬츠	자전거	
u001	2	5	4	1	필터
u002	1	3	5	0	
u003	2	4	5	1	
u004	4	4	2	2	
u005	1	2	3	1	

product_id	서울	경기	부산	대구	울산	
G1142950275	1	3	1	1	1	지역
G1144898993	0	3	2	1	0	
G1144919838	1	1	1	0	1	
G1144920123	2	3	3	2	3	
G1144920813	1	1	1	1	1	
G1144920958	2	3	1	2	2	
G1144977049	0	1	0	0	1	
G1144982391	1	0	1	1	0	

product_id	20대	30대	40대	50대	
G1142950275	0	2	3	4	나이
G1144898993	1	1	1	3	
G1144919838	0	2	1	1	
G1144920123	2	4	3	4	
G1144920813	1	1	2	1	
G1144920958	2	2	3	3	
G1144977049	1	1	0	0	
G1144982391	0	1	1	1	

(그림 10) 세분화된 데이터베이스 결과

값을 증가한다. 위 알고리즘을 적용한 모 전자상거래 업체의 품목별, 지역별, 나이별 트랜잭션 데이터에 의한 세분화된 분석 결과 데이터베이스는 아래 (그림 10)과 같은 결과로써 볼 수가 있다.

4. 실험 및 적용결과

4.1 구현 환경

본 시스템은 서버의 하드웨어 사양으로 CPU가 2.4Ghz 펜티엄 IV PC 기반의 Windows 2003 서버이며, RAM 512MB, 웹서버로 IIS Web Server, 데이터베이스는 MS-SQL Server 2000을 사용하였다. 운영자 관리 툴은 Visual Studio.NET C# 으로 개발을 하였으며, XML 문서의 필터링을 위하여 .NET Framework의 XML class인 XMLTextWriter class와 XML TextReader class를 이용하였으며, 적용 코드는 아래 (그림 11)과 같다. 사용자 프로파일에 의한 상품 추천과 운영자를 위한 사용자 세분화 분석 결과는 ASP와 컴포넌트인 Chart FX.NET을 이용하였다.

```

namespace CodeBank
{
    public class XmlFilter : System.Web.UI.Page
    {
        protected System.Web.UI.WebControls.TextBox txtOriginal;
        protected System.Web.UI.WebControls.TextBox txtFiltered;

        private void Page_Load(object sender, System.EventArgs e)
        {
            int orderIDMin = 10900;
            string path = Server.MapPath("Orders.xml");
            XmlTextReader sReader = null;

            try
            {
                //Write out original XML document
                sReader = new XmlTextReader(path);
                txtOriginal.Text = sReader.ReadToEnd();
            }
            finally
            {
                sReader.Close();
            }

            XmlTextReader reader = null;
            XmlTextWriter writer = null;

            try
            {
                //Create XmlTextReader
                reader = new XmlTextReader(path);
                StringWriter sw = new StringWriter();
                writer = new XmlTextWriter(sw);
                writer.Formatting = Formatting.Indented;
                writer.Indentation = 4;
                //Add root tag for filtered XML document
                writer.WriteStartElement("root");
                foreach (XmlNode node in reader.Nodes["Orders"])
                {
                    while (reader.Read())
                    {
                        if (reader.Name.Equals("order") && reader.NodeType == XmlNodeType.EndElement)
                        {
                            reader.MoveToAttribute("OrderID");
                            //Compute OrderID attribute against orderIDMin integer value
                            if (int.Parse(reader.Value.ToString()) >= orderIDMin)
                            {
                                reader.MoveToElement();
                                //Write desired node to the filtered document XML
                                writer.WriteNode(reader,true);
                            }
                        }
                    }
                }
                writer.WriteEndElement();
                txtFiltered.Text = sw.ToString();
            }
            finally
            {
                writer.Close();
                reader.Close();
            }
        }
    }
}
    
```

(그림 11) XML Class를 이용한 XML 필터링 소스코드

4.2 적용 데이터

본 논문의 운영자 및 사용자 프로파일 적용결과를 명확히 비교, 검증하기 위하여 현재 운영 중인 레저상품 전자상거래 사이트인 GOALPINE[16]의 상품 데이터를 이용하였다. 상품의 목록은 <표 4>와 같이 등산 및 자전거 상품으로 대분류할

<표 4> 상품 분류 카테고리

구분	제조원	품목
등산	A 제조원	셔츠(2), 팬츠(2)
	B 제조원	자켓(4), 셔츠(1), 베스트(5), 팬츠(2), 중형배낭(2), 모자(2), 장갑(2)
자전거	C 제조원	C 제조원(50)
	D 제조원	D 제조원(32)

수 있고, 세부적으로 제조원과 품목에 따라 1개~7개까지의 카테고리로 다시 분류된다. 단, 제조원은 익명인 A, B, C, D 제조원으로 하며, 품목명 괄호 내의 숫자는 GOALPINE에서 제공한 품목의 개수이다.

4.3 운영자 프로파일을 적용한 세분화 결과

전자상거래 운영자가 고객 세분화 분석 결과 및 사용자 프로파일에 적용하기 위해서는 운영자 프로파일을 생성하여야 한다. 본 시스템에 적용할 운영자 프로파일은 관심품목별, 지역별, 나이별로 하여 아래 <표 5>와 같이 구성한다. 이러한 운영자 프로파일은 구매가 발생할 경우 세분화 데이터베이스 및 사용자 프로파일을 구성함에 있어 정보 필터링 역할을 하게 된다.

<표 5> 운영자 프로파일의 구성

운영자 프로파일 ID	운영자 프로파일 설명	프로파일 요소	적용범위
P_1	관심품목	<user_id><name> <order_date> <product_id><Product_catalog>	special
P_2	지역	<product_id><zip_code>	public
P_3	나이	<product_id><age>	public

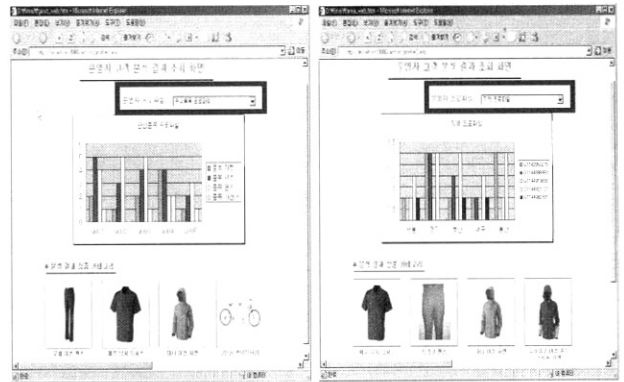
아래 세 개의 표현은 <표 5>에서 생성된 운영자 프로파일의 구성을 기술한 것이다.

$$P_1 = \{P_{1_관심품목}, \{<name>, <order_date>, <product_id>, <product_catalog>\}\}$$

$$P_2 = \{P_{2_지역}, \{<product_id>, <zip_code>\}\}$$

$$P_3 = \{P_{3_나이}, \{<product_id>, <age>\}\}$$

이러한 운영자 프로파일은 상품을 구매하는 사용자들의 트랜잭션 데이터에서 해당되는 항목만 필터링하여 데이터베이스에 저장할 한다. 세분화된 결과로 위 운영자가 설정한 프로파일에 의해서 구성되며 (그림 12)는 운영자 프로파일 중에서도 관심품목과 지역에 따른 세분화된 분석 결과를 보여주고 있으며, 상단에 운영자 프로파일을 선택할 수 있는 목록(사각형 표시 부분)이 있다. 운영자의 선택된 프로파일에 의해서

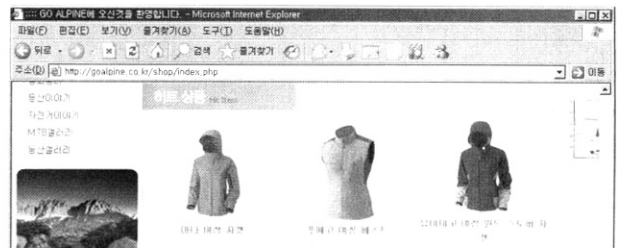


(그림 12) 관심품목, 지역에 따른 세분화 분석 결과

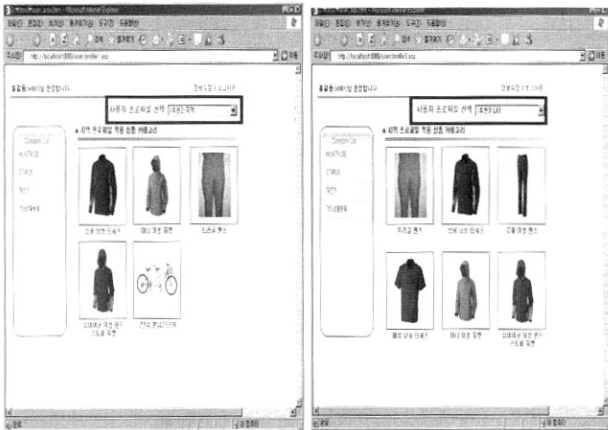
분석 차트 및 결과 상품 목록을 보여 주게 된다. 운영자는 이러한 분석 결과를 토대로 상품의 판매 및 배치 전략을 다시 세울 수 있다.

4.4 사용자 프로파일을 적용한 개인화 결과

해당 프로파일에 적용할 사용자의 예로 “상품을 구매한 이력이 있는 울산 거주자, 나이는 42세, u001의 아이디 사용자” 이라고 한다. u001 사용자는 구매한 이력이 있기 때문에 <표 5>의 운영자 프로파일의 적용을 받아 P_1, P_2, P_3 의 사용자 프로파일을 가지게 된다. 만약 u001 사용자가 구매가 없다면 P_2, P_3 의 사용자 프로파일만 가지게 된다. 생성된 프로파일은 다양한 상품을 추천 받을 수 있는 선택 목록으로 구성이 된다. 따라서 u001 사용자가 로그인 후 프로파일 목록을 어떤 것을 선택하는지에 따라 개인화 된 추천 상품의 목록은 새로이 갱신된다. (그림 13)은 현재 GOALPINE 사이트에서 적용 중인 협업 필터링과 판매량이 많은 상품에 대한 추천화면이다. 모든 로그인 사용자는 추천하는 상품이 모두 동일하며 개인 취향에 맞는 어떠한 상품 추천을 받을 수가 없다. 반면에 (그림 14)는 생성 프로파일 중에서 지역과 나이 프로파일을 적용한 상품 추천화면이다. 지역과 나이 프로파일은 화면 상단에 선택할 수 있는 목록(사각형 마크부분)이 있어 선택에 따른 다양한 상품 추천을 받을 수가 있다. 또한 사용자 초기 추천 화면은 여러 개의 프로파일 중에서 사용자가 주로 이용하는 사용자 프로파일의 빈도수를 계산하여 높은 빈도수의 값을 초기 추천 화면의 사용자 프로파일로 결정한다.



(그림 13) GOALPINE의 기존 프로파일 적용 상품 추천화면



(그림 14) 여러 형태 사용자 프로파일 적용 상품 추천화면 (지역, 나이)

4.5 성능 평가

제안된 방법의 유용성을 검증하기 위해서 고객 50명과 <표 4>에서 분류한 상품 데이터에 대해서 예측 실험을 수행하였다. 추천 효율을 측정하는 방법은 정확도를 기준으로 시계열 분석 방법인 MAE(Mean Absolute Error : 평균 절대 오차)와 ROC(Receiver Operating Characteristic : 반응자 작용 특성)를 사용하였다[15]. MAE는 예측 값과 실제 값의 절대 오차 평균한 값으로써 최소화되어야 정확도를 높다고 할 수 있다. 또한 ROC는 의사결정 지위의 정확도를 측정하는 방법의 하나로 사용자가 가장 양질의 아이템을 선택할 수 있도록 어떻게 효율적으로 예측하는지 여부를 평가하는 방법이다. ROC는 Sensitivity(민감도), Specificity(특이도), Accuracy(정확도), Error rate(오차율)에 의해 평가할 수 있다. ROC를 사용할 경우 정확도 분석을 위해 good과 bad 콘텐츠로 구분하는데, 평가값(rating)이 1~5사이의 값을 가지는 콘텐츠 중에서 rating이 4, 5의 값을 good(positive) 콘텐츠라 하고 1, 2, 3의 값을 bad(negative) 콘텐츠라고 정의한다. true는 예측 값이 올바르게 평가된 것을 의미하고, false는 예측 값이 잘못 평가된 것을 의미한다.

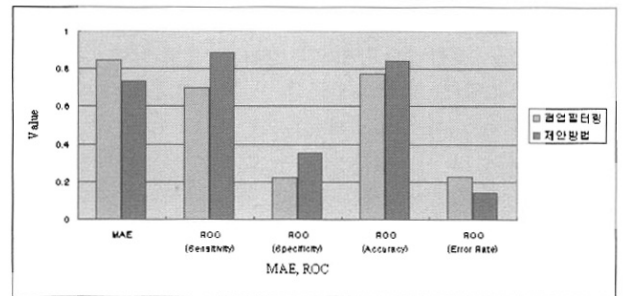
- Sensitivity는 임의로 선택된 평가 값이 4 이상인 콘텐츠가 추천된 확률이고 다음 식과 같이 정의된다.

$$\text{Sensitivity} = \text{true positive} / \text{total actual positive}$$
- Specificity는 임의로 선택된 평가 값이 3 이하인 콘텐츠가 추천되지 않을 확률이고 다음 식과 같이 정의된다.

$$\text{Specificity} = \text{false negative} / \text{total actual negative}$$
- Accuracy는 추천의 정확도(예측이 맞은 경우)를 의미하며 다음 식과 같이 정의된다.

$$\text{Accuracy} = (\text{true negative} + \text{true positive}) / \text{grand total}$$
- Error rate는 추천의 오차율(예측이 틀린 경우)을 의미하여 다음 식과 같이 정의된다.

$$\text{Error rate} = (\text{false negative} + \text{false positive}) / \text{grand total}$$



(그림 15)는 본 논문에서 제안한 여러 형태의 사용자 프로파일 방법과 기존 GOALPINE 사이트에서 적용 중인 협동적 필터링 기반의 프로파일 방법에 대한 MAE 및 ROC 측정값을 비교한 결과 값을 나타내고 있다. 특히 측정에서, Accuracy를 위하여 하나의 항목으로 구성된 프로파일을 이용하는 것이 아니라 여러 항목을 통합적으로 이용함으로써 유사한 상품 선호 성향을 가진 사용자들을 그룹화하고 상품을 추천한다. Sensitivity는 구매자와 같은 나이 대, 같은 지역, 결제 방법 등 세 가지 조건을 모두 만족하는 다른 사용자에게 상품을 추천하기 때문에 높은 값을 가진다.

위 결과 값을 볼 때 사용자들의 성향을 분석하기 위해서 별도의 솔루션이나 모듈을 운영하거나 추가하지 않고, 운영자가 직접 분석 프로파일을 정의함으로써 성향 분석이 용이하고, 트랜잭션 기반으로 사용자 추천이 이루어지기 때문에 사용자가 입력한 부정확한 정보를 이용하는 것보다 더 정확한 사용자 추천이 이루어진다.

5. 결론

기존의 사용자 프로파일은 전자상거래를 이용하는 사용자의 관심도나 사용자의 성향을 분석하기에는 많은 비용과 어려움이 있을 뿐만 아니라 전자상거래를 이용하는 사용자들의 기본적인 정보는 형식적이기 때문에 다양한 프로파일을 생성할 수가 없었다. 본 논문에서는 전자상거래를 이용하는 사용자들의 상품에 구매 성향을 잘 파악할 수 있는 트랜잭션 데이터 기반에 사용자 기본정보를 추가하여 필요한 정보를 필터링함으로써 사용자 프로파일을 여러 형태로 생성할 수가 있었다. 해당 트랜잭션 데이터와 사용자 기본정보를 필터링을 위하여 먼저 운영자는 운영자 관리 툴을 이용하여 운영자 프로파일을 생성하였으며, 이 운영자 프로파일은 로그인 한 사용자들의 행위 결과인 트랜잭션 데이터를 필터링하여 또 다른 형태의 사용자 프로파일로 재구성되며, 사용자 프로파일은 운영자에 의해서 만들어진 프로파일에서 파생되었기 때문에 다양한 운영자 프로파일은 다양한 사용자 프로파일로 재구성되어진다. 그리고 CRM과 eCRM을 지원하기 위한 기존의 시스템은 하나의 복잡한 모듈이나 패키지로서 제공되고 개별 시스템에 적합하도록 하기 위해서는 별도의 작업을 통하여 시스템을 재구성한 후 제공하여야 하는 문제점을 가지고 있

었다. 이러한 문제점을 해결하기 위해서 운영자가 직접 분석할 프로파일을 생성하였으며, 이 결과로 운영자에게 분석에 필요한 프로파일에 의한 세분화된 자료를 제공할 수가 있었다. 추가적으로 사이트 중심의 고정된 프로파일에 의한 상품 추천이 아니라 각 사용자가 가지는 다양한 프로파일에 의한 상품추천을 해 주기 때문에 선택된 프로파일 분류에 의해서 각각 다른 추천을 제공해 주는 이점 또한 가진다. 최종적으로 본 논문의 프로파일을 적용하고 성능을 입증하기 위하여 현재 운영 중이며 협업기반 프로파일을 적용중인 GOALPINE 사이트와의 성능비교를 통하여 효율성을 입증하였으며, 더불어 GOALPINE에서 제공된 일부 상품을 토대로 본 논문의 프로파일을 적용한 예시 웹사이트를 통하여 결과를 살펴보았다.

향후 과제로 실제 GOALPINE 사이트의 전반적인 모든 부분에 본 논문의 사용자 프로파일을 적용 구축할 계획이며, 추가적으로 사용자 프로파일을 다양하게 활용하여 적당한 사용자에게 상품의 추천 내용을 선택배포 할 수 있는 SDI 시스템을 추가한 확장된 운영자 관리 툴을 연구 중에 있다.

참 고 문 헌

[1] Alex Berson, Stephen Smith, Kurt Thearing, "Building Data Mining Application for CRM", McGraw-Hill, pp.4-14, ISBN 0-07-134444-6

[2] G. Rossi, D. Schwabe, D., Guimaraes, R. M., "Designing Personalized Web Applications", Prof. of the International World Wide Web Conference (WWW), Amsterdam, 2001.

[3] B. Sarwar, G. Karypis, J. Konstan, and J.Riedl, "Item-based Collaborative Filtering Recommendation Algorithms", Proc. of the Tenth International World Wide Web Conference on WWW, pp.285-295, 2001.

[4] Ranjit Bose, "Customer relationship management: key components for IT success" Industrial Management & Data Systems, Vol.102, No.2 (March 1, 2002)

[5] Ugru Cetintemel, Michael J. Franklin, and C. Lee Giles, "Self-Adaptive User Profiles for Large-Scale Data Delivery", Proceedings of the International Conference on Data Engineering, San Diego, CA, pp.622-633, February, 2000.

[6] R.Cooley, Mobasher, Bamshad, H.Srivastava, "Data Preparation for Mining World Wide Web Browsing Pattern", The Journal of Knowledge and Information System, Vol.1, No.1, 1999.

[7] B.Krulwich, "Lifestyle Finder : Intelligent User Profiling Using Large-Scale Demographic Data", Artificial Intelligent Magazine Vol.18, No.2, pp.37-45, 1997.

[8] Mehmet Altinel, Michael J. Franklin, "Effective Filtering of XML Document for Selective Dissemination of Information", Proceedings of the 26th VLDB Conference, Cairo, Egypt, 2000.

[9] Chee Yong Chan, Pascal Felber, Minos N. Garofalakis, and Rajeev Rastogi. Efficient filtering of XML documents with XPath expressions. in ICDE 2002.

[10] Y. Diao, M. Altinel, M. J. Franklin, H. Zhang, and P. Fischer. Path sharing and predicate evaluation for high-performance XML filtering. ACM Transactions on Database Systems, 28(4):467-516, 2003.

[11] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering", Artificial Intelligent

Review, pp.394-408, 1999.

[12] M.Claypool, A.Gokhale, T.Mirande, P.Murnikov, D.Netes and M.Sarti, "Combining Content-Based and Collaborative Filters in an Online Newspaper", ACM SIGIR Workshop on Recommender System, Berkeley, CA, 1999.

[13] Rachael Rafter, Keith Bradley, Bary Smyth, "Personalised Retrieval for Online Recruitment Services", In Proceedings of the 22nd Annual Colloquium on Information Retrieval (BCS-IRSG 2000), Sidney Sussex Collage, Cambridge, UK, 2000.

[14] G.Adomavicius and A.Tuzhilin, "Using Data Mining Methods to build Customer Profiles", IEEE, February, pp.74-82, 2001.

[15] N.Good, B.Schafer, J.Konsta, A.Borchers, B.Sarwar, J.Herlocker, J.Riedle, "Combining Collaborative Filtering with Personal Agents for Better Recommendation", Proc. of the AAAI conference, pp.439-446, 1999.

[16] http://www.goalpine.co.kr/shop_rocomsystem

[17] Balabanovic, Marko, and Yoav Shoham. "Fab : Content-based, Collaborative Recommendation". Communications of the ACM, 40(3):66-72, 1997.

[18] Dunja Mladenic, Personal WebWatcher: Implementation and Design, Technical Report IJS-DP-7472, October, 1996.

[19] Bruce Krulwich, Cbad Burkey "The InfoFinder Agent: Learning User Interest through Heuristic Phrase Extraction," AgentSoft Ltd, Andersen Consulting LLP, 1995.

[20] Bruce Krulwich, "Learning document category description through the extraction of semantically significant phrase," Center for Strategic Technology Research Andersen Consulting LLP 100 South Wacker Drive, Chicago, IL 60606, 1995.

고 재 진



e-mail : jjkoh@mail.ulsan.ac.kr

1972년 서울대학교 응용수학과(공학사)

1981년 서울대학교 대학원 계산통계학과 (이학석사)

1990년 서울대학교 대학원 컴퓨터공학과 (공학박사)

1975년~1979년 한국후지쓰(주) 기술개발부 사원

1979년~현재 울산대학교 컴퓨터정보통신공학부 교수

관심분야: DB시스템, 전문가 시스템, DB설계, ERP

안 형 근



e-mail : hkahn@mail.ulsan.ac.kr

2000년 한국방송통신대학교 컴퓨터과학과 (이학사)

2003년 울산대학교 정보통신대학원 정보통신공학과(공학석사)

2006년 울산대학교 대학원 컴퓨터정보통신공학부(박사과정 수료)

1997년~2004년 현대오토시스템 기술지원부

2004년~현재 (주)CFIC 기업부설연구소 연구소장

관심분야: 멀티미디어DB, DB설계/분석, ERP, BPM, Workflow