

고객의 선호 특성 정보를 이용한 상품 추천 시스템

성 경 상* · 박 연 출** · 안 재 명*** · 오 해 석****

요 약

전자상거래 시스템의 보급이 활성화되기 시작하면서, 사용자의 필요와 욕구에 밀착한 적응형 전자상거래 에이전트의 필요성이 증대되고 있다. 이와 같은 적응형 전자상거래 에이전트는 사용자의 행위를 모니터링하고 자동 분류하여 사용자의 취향을 학습하는 기능을 요하게 되었다. 이러한 기능을 가지는 적응형 전자상거래 에이전트를 구축하기 위해서, 본 논문에서는 사용자 개인의 관심정보와 선호하는 상품에 대한 호감도를 고려한 적응형 전자 상거래 에이전트 시스템을 제안한다. 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 제공 할 수 있도록 한다. 본 시스템에서는 모니터링을 통하여 사용자 의도를 파악하는 모니터 에이전트, 사용자의 행동성향을 학습 한 후 행동 패턴이 유사한 그룹을 참조하는 유사도 참조 에이전트, 사용자의 행위의 변화에 따른 개인화된 행동 DB를 구축할 수 있는 관심 추출 에이전트로 구성하였다.

Goods Recommendation System using a Customer's Preference Features Information

Kyung-Sang Sung* · Yeon-Chool Park** · Jae-Myung Ahn*** · Hae-Seok Oh****

ABSTRACT

As electronic commerce systems have been widely used, the necessity of adaptive e-commerce agent systems has been increased. These kinds of adaptive e-commerce agents can monitor customer's behaviors and cluster them in similar categories, and include user's preference from each category. In order to implement our adaptive e-commerce agent system, in this paper, we propose an adaptive e-commerce agent systems consider customer's information of interest and goodwill ratio about preference goods. Proposed system build 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. The proposed system composed with three parts ; Monitor Agent which grasps user's intension using monitoring, similarity reference Agent which refers to similar group of behavior pattern after learned behavior pattern of user, Interest Analyzing Agent which personalized behavior DB as a change of user's behavior.

키워드 : 적응형 전자상거래 에이전트(Adaptive Electronic Commerce Agent), 고객 프로파일(User Profile), 귀납적 기계학습(Inductive Machine Learning)

1. 서 론

쇼핑몰 등에서 자신에게 맞는 상품들을 추천받는 서비스는 이미 미국 시장에서는 어느 정도 일반적인 일이 되었으며, 국내에서도 이미 그 자리가 다져지고 있다. 철저하게 고객의 필요와 욕구에 밀착된 인터넷 서비스를 제공한다는 개인화(Personalize)의 취지에 비추어 볼 때, 이와 같은 추천 시스템은 필수적으로 입지를 굳힐 것으로 전망된다. 이런 상황이라면 고객에게 어느 정도의 개인화 서비스를 제공하느냐가 앞으로 쇼핑몰에 대한 성패를 좌우하게 될 것임에 틀림없다.

최근에는 데이터마이닝(DataMining)이나 CRM(Customer Relationship Management) 시스템 등을 이용하여 고객을 유치, 관리 및 서비스를 하고 있다. 하지만, 이러한 시스템을 개발하고 관리하는데 있어서는 과도한 비용과 시간이 소요되는 점과 관리자가 시스템을 올바르게 이해하고 관리하는데 있어 많은 어려움이 있다. 이를 보완하기 위해 좀 더 쉽고 이용이 간편하며 비용이 적게 드는 소규모 개인화 서비스에 대한 연구가 활발하게 진행되고 있다.

개인화 서비스란, 무분별하고 과도하게 제공되는 정보들에 대해 여과 기술을 이용하여 무분별하게 제공되어지는 정보들을 정제(精製)함으로써 사용자 개인의 취향에 알맞은 정보를 제공하는 서비스를 말한다. 이러한 최근의 개인화 서비스는 정보의 추천을 통해 제공되고 있으나, 추천의 정확도에 있어서 한계를 보이고 있다는 문제점이 드러나고 있다.

* 준 회원 : 경원대학교 대학원 컴퓨터학과
 ** 준 회원 : 숭실대학교 대학원 컴퓨터학과
 *** 종신회원 : (주)리테일테크 대표이사
 **** 종신회원 : 경원대학교 소프트웨어대학 교수/부총장
 논문접수 : 2003년 12월 20일, 심사완료 : 2004년 7월 7일

또한 사용자간의 유사도를 고려하는데 있어서 항목간의 중요도, 즉 가중치를 고려하지 않는 문제점을 지니고 있을 뿐만 아니라 공통된 선호도 항목이 적을 경우 발생하는 희소성(sparsity)문제와 사용자의 수가 증가함에 따라 수행속도가 저하되는 확장성(scalability)문제로 인하여 효율성이 떨어지는 문제점들을 지니고 있다. 이러한 기존 방식의 문제점들을 어느 정도 해결하고자 새로운 추천 기법을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 관련연구에 대해 살펴보고, 제 3장에서는 연구 목적 및 방법과 시스템의 전체적인 구조와 각 모듈들에 대해 다룬다. 제 4장에서는 제안하는 시스템의 선호도 추출에 관련된 알고리즘에 대해 살펴보고, 제 5장에서는 제안하는 시스템의 구현 환경과 추천되어져 가는 과정에 대해 살펴보고, 국내·외에서 현재 인공지능기능을 탑재한 개인화된 시스템을 운용중인 웹 쇼핑몰과 성능평가를 통해 제안하는 시스템의 우수성을 살펴본 후, 제 6장에서는 결론과 향후 연구 과제를 통해 제안하는 시스템이 적용되어질 수 있는 방향을 제시하도록 기술한다.

2. 관련 연구

2.1 WiseWire 지능형 에이전트

WiseWire 엔진은 WiseWire사에서 만든 지능형 에이전트로, 인터넷에서 정보를 검색할 때 사용자의 기존 검색 패턴을 분석한 후 얻은 프로파일을 이용하여 사용자들에게 각 개인이 원하는 정보를 보다 쉽게 검색할 수 있도록 하는 기능들을 제공한다.

WiseWire 엔진은 관심정보와 비 관심정보를 기반으로 사용자의 관심(Profile)을 기계학습을 이용하여 학습하고, 이렇게 학습된 사용자의 관심을 이용하여 사용자 적응력 있는 검색을 수행한다. 또한, WiseWire 엔진은 각 사용자에 대한 관심 학습뿐 아니라, 일군의 그룹에 대한 협업적 학습(collaborative learning)을 수행한다[5].

2.2 오이 뮤직의 개인화 전략(<http://www.oi.co.kr>)

오이 뮤직은 현재 인공지능 개념을 도입하여 앨범을 추천해주는 웹 사이트로, 회원 가입 과정을 거쳐 로그인을 하게 되면 인공지능 추천 센터를 통하여 사용자의 나이와 성별에 의해, 일차적인 추천이 이루어진다. 또한 고객들의 구매 정보에 따라 앨범에 대한 추천도 이루어지게 된다. 자신의 음악스타일에 따라 앨범 평가를 하는 과정을 거치게 되는 동안 개인화 서비스를 받을 수 있게 된다. 또한 앨범을 구매한 다른 사람들의 구매 경향도 알 수 있어 구매에 도움을 받을 수 있다[13].

2.3 그룹화를 이용한 에이전트 기반의 전자도서관 시스템

에이전트 기반 전자도서관 시스템은 웹을 통해 도서정보

를 검색하는 사용자들의 도서 정보 관심도를 모니터 에이전트에 의해 추출하게 되고 해당 정보를 대상으로 학습 과정을 거쳐 사용자별 프로파일을 구축한다. 이 과정에서 모니터 에이전트에서 추출되는 사용자 히스토리와 학습 예제 정보는 사용자가 행하는 다양한 작업의 결과를 이용한 차별화된 중요도가 적용된다. 학습 에이전트를 통해 구축된 사용자 관심 정보는 각 사용자에게 세부 DB 추천 모듈과 관심 정보를 푸쉬해 주는 에이전트에 적용된다. 또한 전자도서관에 가입하는 사용자들의 공통된 그룹 정보를 관리하여 이를 통해 동일 그룹내의 사용자들 사이의 정보 이용이 가능하게 한다[6, 11].

이와 같이 WiseWire에서 제공하는 웹 엔진을 이용할 경우 공용의 시스템에서는 개인화된 정보를 정확히 제공할 수 없으며, 오이 뮤직에서 제공하는 인공지능화 된 시스템에서는 사용자가 선호하는 상품에 대한 추천 여부에 대해서도 한정적인 정보만을 이용하여 일차적인 추천만을 하게 되므로 보다 정확한 추천을 하기에는 그 정도가 매우 미비한 실정이었다. 이러한 문제점들을 해결하기 위한 방안으로는 사용자의 정보를 서버에서 관리하며, 개인화된 정보를 유지할 수 있도록 하며, 정확한 추천을 기하기 위해 실시간적으로 적용되어지는 사용자의 정보를 통하여 보다 다양한 정보를 활용하여 개인화된 서비스를 받을 수 있도록 한다.

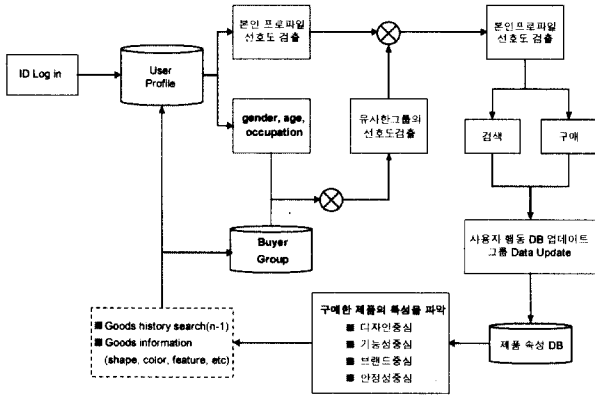
3. 프로파일링 에이전트의 설계

본 논문에서는 사용자의 행위에 따라서 적응형으로 동작하는 지능형 전자상거래 에이전트 구축을 위한 방안을 제시한다. 사용자의 행위를 포괄적으로 모니터링하여 사용자의 의도를 파악하는 모니터 에이전트, 사용자의 행동성향을 파악하여 행동 패턴이 유사한 그룹을 참조하는 유사도 참조 에이전트, 사용자의 행위의 변화에 따른 개인화된 행동 DB를 구축할 수 있는 관심 추출 에이전트로 구성하였다. 적응형 전자상거래 에이전트 시스템의 설계 목적은 사용자의 의도를 파악하여, 사용자가 관심을 가지는 정보를 추출하여 생성하는 사용자 프로파일의 정확도에 있다. 본 논문에서는 정확한 사용자 프로파일을 학습하고, 이를 기반으로 적응형 서비스를 수행할 수 있도록 다음과 같은 전자상거래 에이전트 구조를 제안한다.

3.1 시스템 구조

(그림 1)은 선호 특성 추출을 위한 고객 프로파일링 시스템의 전체적인 구성도로서 학습 에이전트를 이용하여 사용자에 대한 선호도를 추출한다. 추출되어진 상품에 대한 사용자의 구매 성향과 성향별로 그룹화 된 고객단의 구매 성향을 분석하여 선호도를 계산하게 되며, 추출되어진 선호도를 이용하여 각 상품에 대한 가중치를 고려하여 사용자에

계 가장 적합한 상품을 추천해 줄 수 있도록 한다.



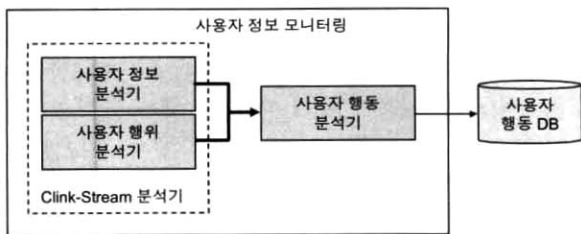
(그림 1) 전체적인 시스템 구성도

사용자의 행위를 모니터하여 각 사용자가 브라우징한 상품, 구매한 상품 등을 실시간으로 DB를 갱신하여 각 정보는 사용자가 전자상거래 시스템을 브라우징하면서 수행한 행위들을 정확히 기록하는 기능을 수행한다. 유사도 참조 에이전트는 모니터 에이전트에서 생성한 사용자의 행위로부터 유사한 관심 분야별로 그룹을 만드는 기능을 수행한다. 사용자 행동 성향을 파악한 후 귀납적 기계학습 방식을 적용하여 사용자의 취향을 학습하는 기능을 수행하는데 있어 사용자 선호도 추출을 위한 관심 컴포넌트를 적용한다.

3.2 구매 성향 파악을 위한 모니터 에이전트

적용형 전자 상거래 시스템을 위해서 가장 먼저 수행 되어야 하는 것은 사용자가 취하는 행동들을 포괄적으로 탐지하는 것이다[10].

이를 위해 모니터 에이전트는 전자상거래를 이용하는 사용자의 행위를 포괄적으로 모니터하여, 사용자의 행위를 추상화한다.



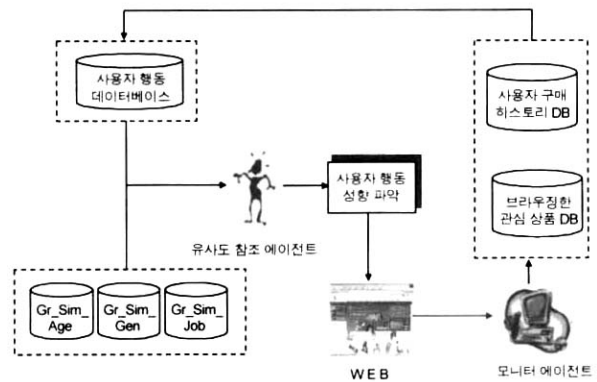
(그림 2) 모니터 에이전트

모니터 에이전트는 사용자의 행위를 모니터하고, 각 사용자가 브라우징한 상품, 구매한 상품 등에 대해 데이터베이스에 실시간적으로 적용되어지도록 한 후, 일련의 작업을 통해 사용자의 행위를 분석한 후 활용하도록 하는데 목적이 있다. 이렇게 모니터 된 정보는 사용자별로 프로파일을 만들기 위해서 사용자별로 나눈 후, 각각의 모니터 된 정보

를 사용자 행위로 추상화하는 과정을 수행한다. 예를 들어, 사용자가 특정 상품 URL에 방문했다는 정보를 특정 상품과, 그 URL에 대한 관심 정보로 추상화 한다.

3.3 유사도 참조 에이전트를 이용한 구매 성향 파악

웹 검색을 통해 구축되어진 사용자 행동 데이터베이스와 유사한 프로파일을 지닌 그룹 내의 행동 지침들을 참조하여 유사도 참조 에이전트를 통해 사용자의 구매 성향을 파악하게 된다.

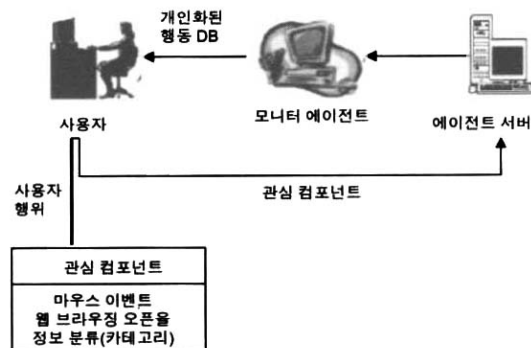


(그림 3) 유사도 참조 에이전트를 이용한 사용자 행동 성향 파악

이러한 과정에 있어 아이템 검색과정과 구매과정은 모니터 에이전트를 통해 수집되어지고, 사용자 행위에 관련된 정보는 사용자 구매 히스토리 데이터베이스와 브라우징한 관심 상품 데이터베이스에 실시간적으로 저장되어 기록 한다.

3.4 고객관심도 측정 및 생성

마우스 이벤트 효과나 브라우징 오픈에 의해 발생되어지는 사용자 행위에 따른 에이전트 서버에 사용자의 관심 여부를 전송하게 된다. 이로 인해 에이전트 서버는 사용자 행위에 의한 정보를 획득하게 되고, 이를 바탕으로 사용자 선호도를 추출하게 된다.



(그림 4) 사용자 행위에 의한 관심 컴포넌트

이러한 사용자 행위에서 발생되어지는 잠재적인 관심정보를 관심 컴포넌트라고 정의하며, (그림 4)는 사용자 행위에 의해 나타나는 관심 컴포넌트가 구축되어지는 과정을 보여주고 있다.

4. 선호도 추출 알고리즘

4.1 사용자의 선호도 추출 알고리즘

$$P_{u,j} = \frac{P_{u,item[j]}}{\sum_{i=1}^n P_{u,item[i]}} * W_p + S_{u,item[j]} \quad (1)$$

$P_{u,j}$ 는 사용자 u 의 상품(item) j 에 대한 선호도를 나타내며, $P_{u,item[j]}$ 는 사용자 u 가 상품 j 를 구매한 수를 나타낸다. 그리고 $P_{u,item[i]}$ 는 동일 카테고리 내에서 사용자 u 가 기존에 구매했었던 상품의 수를 나타낸다. n 은 전체적인 아이템 수를 나타낸 것이고, W_s 는 상품 검색에 따른 가중치를 가리키며, W_p 는 상품 구매에 따른 가중치를 나타내는 것으로 가중치에 따른 범위는 '0~1' 사이의 피어슨 계수를 이용하였다. $S_{u,item[j]}$ 는 사용자 u 의 상품 j 에 대한 검색율을 나타내고, $S_{u,item[i]}$ 는 동일 카테고리 내에서 사용자 u 가 기존에 검색했었던 상품의 검색율을 가리킨다.

4.2 형성되어진 그룹에 대한 선호도 추출 알고리즘

식 (2)는 사용자와 유사한 연령 그룹에 따른 선호도 추출 알고리즘을 표현하였으며, 식 (3)은 동일한 성별 그룹에 따른 선호도 추출 알고리즘, 식 (4)는 유사한 직업 그룹에 따른 선호도 추출 알고리즘을 나타낸다.

가중치는 사용자에 대한 구매와 검색 이벤트가 발생 할 때마다 모든 카테고리 내에서 전체 상품에 대한 구매율과 검색율을 계산하여 구매 가중치와 검색 가중치에 대한 대비율로 실시간적으로 계산되어진다.

$$GA_{u,a[j]} = \sum_{k=1}^m \left(\frac{gA_{-P_{u,a[j]}}}{\sum_{i=1}^n gA_{-P_{u,a[i]}}} \right)_{[k]} * W_{pa} + \sum_{k=1}^m \left(\frac{gA_{-S_{u,a[j]}}}{\sum_{i=0}^n gA_{-S_{u,a[i]}}} \right)_{[k]} * W_{sa} \quad (2)$$

식 (2)에서 $GA_{u,a[j]}$ 는 사용자 u 와 연령이 유사한 사용자 a 사이의 상품(item) j 에 대한 선호도를 나타낸 것으로, $gA_{-P_{u,a[i]}}$ 과 $gA_{-S_{u,a[i]}}$ 는 동일 카테고리 내에서 사용자 u 와 연령이 유사한 사용자 a 가 기존에 구매했었던 상품의 수와 검색율을 가리킨다. m 은 사용자 수, n 은 아이템 수 그리고 W_{pa} 과 W_{sa} 는 상품 구매와 검색에 따른 연령에 부과되는 가중치를 가리킨다.

$$GS_{u,a[j]} = \sum_{k=1}^m \left(\frac{gS_{-P_{u,a[j]}}}{\sum_{i=1}^n gS_{-P_{u,a[i]}}} \right)_{[k]} * W_{ps} + \sum_{k=1}^m \left(\frac{gS_{-S_{u,a[j]}}}{\sum_{i=0}^n gS_{-S_{u,a[i]}}} \right)_{[k]} * W_{ss} \quad (3)$$

식 (3)에서 $GS_{u,a[j]}$ 는 사용자 u 와 성별이 같은 사용자 a 사이의 상품(item) j 에 대한 선호도를 나타낸 것으로, $gS_{-P_{u,a[i]}}$ 과 $gS_{-S_{u,a[i]}}$ 는 동일 카테고리 내에서 사용자 u 와 성별이 같은 사용자 a 가 기존에 구매했었던 상품의 수와 검색율을 가리킨다. m 은 사용자 수, n 은 아이템 수 그리고 W_{ps} 과 W_{ss} 는 상품 구매와 검색에 따른 성별에 부과되는 가중치를 가리킨다.

$$GO_{u,a[j]} = \sum_{k=1}^m \left(\frac{gO_{-P_{u,a[j]}}}{\sum_{i=1}^n gO_{-P_{u,a[i]}}} \right)_{[k]} * W_{po} + \sum_{k=1}^m \left(\frac{gO_{-S_{u,a[j]}}}{\sum_{i=0}^n gO_{-S_{u,a[i]}}} \right)_{[k]} * W_{so} \quad (4)$$

식 (4)에서 $GO_{u,a[j]}$ 는 사용자 u 와 직업이 유사한 사용자 a 사이의 상품(item) j 에 대한 선호도를 나타낸 것으로, $gO_{-P_{u,a[i]}}$ 과 $gO_{-S_{u,a[i]}}$ 는 동일 카테고리 내에서 사용자 u 와 직업이 유사한 사용자 a 가 기존에 구매했었던 상품의 수와 검색율을 가리킨다. m 은 사용자 수, n 은 아이템 수 그리고 W_{po} 과 W_{so} 는 상품 구매와 검색에 따른 유사한 직업에 부과되는 가중치를 가리킨다.

$$Tot_pre_{[index]} = P_{u,j} + GA_{u,a[j]} + GS_{u,a[j]} + GO_{u,a[j]} \quad (5)$$

식 (5)는 사용자가 선호하는 선호도와 사용자와 유사한 그룹이 갖게 되는 선호도를 가지고 최종적인 아이템을 추천해 주는 알고리즘을 나타낸다.

$Tot_Pre_{[index]}$ 는 $index$ 번째 아이템에 대한 사용자의 선호도를 나타내는 것으로 $P_{u,j}$ 는 사용자 u 에게 아이템 j 를 추천했을 때 예측 결과 값을 나타내고, $GA_{u,a[j]}$ 은 연령이 유사한 사용자 u 와 a 사이의 j 번째 아이템에 대한 예측 결과 값을, 그리고 $GS_{u,a[j]}$ 는 성별이 같은 사용자 u 와 a 사이의 j 번째 아이템에 대한 예측 결과 값 그리고 $GO_{u,a[j]}$ 는 직업군이 유사한 사용자 u 와 사용자 a 사이의 j 번째 아이템에 대한 예측 결과 값을 나타낸다.

4.3 실 데이터를 통한 알고리즘 증명

본 장에서는 일반적인 카테고리 형식의 구조를 지니고, 본 논문에서 제안하는 데이터를 기준으로 알고리즘의 적합성을 증명해나가기로 하겠다. <표 1>에서 보여지는 "가전/통신(001) - TV/비디오(004) - 일반TV(001) - 삼성29인치 제품"에 대한 선호도 계산을 위한 과정을 알아보겠다.

<표 1> 가정/통신 카테고리

가정 / 통신 (001)	생활가전(001)	청소기(001) 정수기(002) 세탁기(003) 에어컨(004)
	카메라/캠코더(002)	디지털(001) 일반(002) 캠코더(003)
	오디오/카세트(003)	CD(001) MP3(002) 오디오(003) 스피커(004)
	TV/비디오(004)	일반TV(001) 완전평면TV(002) VTR(003) DVD(004)

Click-Stream을 통해 선택되어진 제품에 대한 나의 선호도 계산은 구매 제품과 제품 검색에 각각 달리 부여되어진 가중치를 가지고 식 (1)을 통해 선호도가 계산되어진다.

예를 들어 “삼성 29인치 TV”를 39번 검색하고 “LG 29인치 TV”를 21번 검색하면 TV에서의 각각의 제품에 대한 나

의 선호도는 <표 2>에서 보는 바와 같이 각각 0.65와 0.35의 결과를 보이게 된다.

<표 2> 나의 선호도

아이템	각 아이টে에 대한 방문 횟수	선호도값
삼성 29인치	39	0.65
LG 29인치	21	0.35

각 제품에 대한 그룹의 선호도 계산은 식 (2)~식 (4)까지 부여되어진 “성별·연령·직업군”에 부여되어진 선호 특성에 따라 각각 달리 가중치가 적용된다. 즉, 구매하거나 검색한 제품 종류에 따른 각각의 선호도가 다르기 때문에 부과되어지는 가중치의 비율도 달리 구분되어 부과되어진다.

<표 3> 알고리즘에 의한 선호도 값 계산 결과

제품명(item)	형성된 그룹(g _{a,index})								선호도값
	g _{a,1}	g _{a,2}	g _{a,3}	g _{a,4}	g _{a,5}	g _{a,6}	g _{a,7}	g _{a,8}	
LG 29인치	8(14)	3(10)	12(20)	15(35)	0(18)	3(15)	0(0)	0(10)	
각 그룹에 대한 유사도	0.606	0.329	0.594	0.561	0.136	0.41	0	0.113	2.749
각 그룹에 대한 가중치 부여후 최종 선호도 측정	0.606	0.279	0.445	0.392	0.081	0.184	0	0.333	2.32
삼성 29인치	5(10)	7(14)	9(10)	12(25)	2(15)	5(14)	4(20)	8(12)	1.21
각 그룹에 대한 유사도	0.392	0.67	0.394	0.437	0.863	0.588	1	0.886	5.23
각 그룹에 대한 가중치 부여후 최종 선호도 측정	0.392	0.569	0.295	0.305	0.517	0.264	0.55	0.265	3.157

<표 3>에서 보이는 내용은 각 제품에 대해 사용자와 유사한 그룹을 선정하여, 각 그룹이 제품을 구매한 회수와 검색한 회수에 대해 제시한다. 예를 들어 LG 29인치 TV를 g_{a,1} 그룹이 8개를 구입하고 4번의 검색결과가 있었다. 삼성 29인치 TV를 g_{a,1} 그룹 5개를 구입하고 10번의 검색결과가 있었다. 가중치가 부여되어진 알고리즘(식 (1)~식 (4))을 거친 후 식 (5)를 통해 최종적인 선호도 값을 도출해낸다. g_{a,1} 그룹에서의 TV 각각의 제품에 대한 선호도 값은 0.606과 0.392가 된다. 그룹 전체에 있어서의 선호도는 제시된 결과를 통해 “삼성 29인치 TV(3.157)”가 “LG 29인치 TV(2.32)”보다 높은 선호도를 갖는다는 것을 확인할 수 있었다.

각 제품에 대해 <표 3>에서 도출되어진 선호도 결과 값은 식 (6)에 적용하여 선호하는 제품을 추천하게 된다.

For a=0, n, 1

$$RG_{[i]} = \sum_{k=0}^p g_{a,[i]} item_{(index)} \quad (6)$$

$$RL = \text{sort}(RG)$$

예를 들어 LG 29인치 TV에 대한 그룹 전체의 선호도 값은 2.32가 되고, 삼성 29인치 TV에 대한 그룹 전체의 선호도 값은 3.157이 된다. 측정되어진 나의 선호도 값과 나와 유사한 그룹들의 최종적인 선호도 값에 대한 총체적인 결과 값은 (0.65 + 2.32)와 (0.35 + 3.157)이 된다.

그러므로 최종적으로 시스템이 사용자에게 추천해 주는 제품은 그 비율에 의해 “삼성 29인치 TV”가 된다.

5. 실험 및 성능 평가

5.1 구현 환경

본 시스템은 서버의 하드웨어 사양으로 RAM은 256MB이고, CPU가 2.4Ghz 펜티엄 IV PC를 기반으로 하는 시스템에서 Windows 2000 server를 운영체제로, IIS Web Server를 기반으로, MS-SQL server 2000을 데이터베이스로 사용하였다.

인터페이스와 사용자 관심정보 분석을 위한 개발언어로 마이크로소프트사의 ASP 3.0과 JDK 1.4를 사용하였다.

5.2 기본적인 등록정보에 대한 사용 변수 및 근거

한국전산원 정보화 통계 DB 시스템에서 2003년 6월을 기준으로 전자상거래 이용 유무에 대한 조사표로서 연령, 성별, 직업군, 지역별로 크게 구분하여 본 시스템에 적용하도록 한다. 하지만 본 시스템에서는 지역별에 대한 차이가 서울을 기점으로 타 지방과 너무 큰 격차를 보이므로 가중치 부여시 제외시켰으며, 구분되어진 기본정보(연령, 성별, 직업군)를 기점으로 가중치 부여에 일조하도록 하였다[14].

직업에 대한 구분은 통계청에서 제시하는 근거를 가지며,

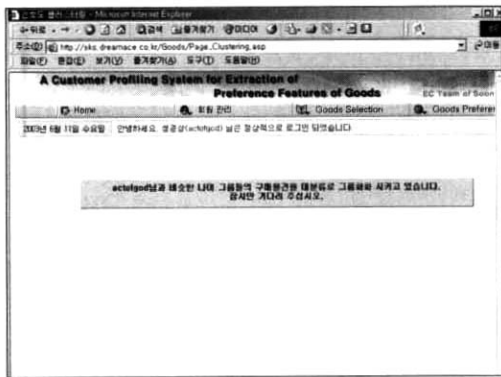
그 구분이 “대-중-소”로 분류되어 있고, 모든 쇼핑몰에서 표준적으로 이용하고 있는 중 분류를 기준으로 대략 15개의 직업군을 기준으로 확정하여 제안하는 시스템에 적용하였다[15].

5.3 시스템 구현 방법

본 시스템은 단지 고객 등록 절차만을 거침으로써 데이터베이스에 등록되어지는 사용자의 기본적인 정보(연령·성별·직업 등)를 이용하여, 사용자와 유사한 정보를 지닌 사람들을 그룹화한다[17]. 형성되어진 그룹에 대해 본 논문에서 제안하는 유사도 측정 알고리즘을 이용하여 각 그룹에 대한 유사도를 분별 측정하여 가중치를 부여한다. 부여되어진 그룹별 가중치를 이용하여 사용자는 웹 쇼핑을 하게 되는데, 중요한 점은 기본적인 정보만을 가지고 제품에 대한 관심정보를 추천받을 수 있다. 즉, 사용자가 구매한 상품이 없을지라도 사용자와 유사한 프로파일을 지니고 있는 그룹에서의 유사 정보를 데이터베이스로부터 추출하게 된다. 이렇게 추출되어진 공통 데이터에 대한 사용자들의 관심에 따른 선호도 비율을 이용하여 사용자는 개별적인 관심 정보를 만들어가게 된다. 이렇게 형성되어진 관심 정보는 웹 쇼핑몰을 이용하게 될 다른 사용자들에게 참조되어 질 수 있도록 유사그룹에 선호도 비율에 적용되는 동시에 사용자 자신에게는 관심 정보 데이터베이스에 추가되어 개별적인 데이터베이스를 형성하게 된다.

5.4 선호 특성 추출을 위한 고객 프로파일링 시스템

선호 특성 추출을 위한 고객 프로파일링 시스템은 로그인 후 함과 동시에 사용자의 프로파일(신상정보, 물품에 대한 관심정보들)을 기반으로 유사한 프로파일을 지니고 있는 사용자 그룹을 참조하여 기존 그룹에 적용하는 단계를 거치게 된다. 그룹이 지니고 있는 해당 물품에 대해 특성들을 파악하여 대분류에 해당되는 카테고리를 사용자에게 보여준다. (그림 5)는 사용자와 유사한 프로파일을 지니고 있는 사용자들을 그룹화하여 유사도를 추출하고 있는 과정을 보여준다.

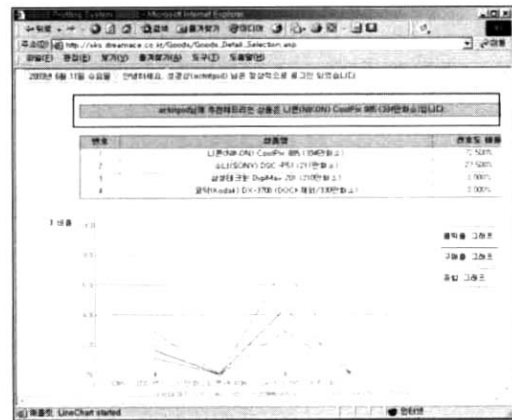


(그림 5) 고객 프로파일을 가지고 선호도 추출 과정

카테고리에 의한 아이템 선택에 있어 개인의 관심도에 따

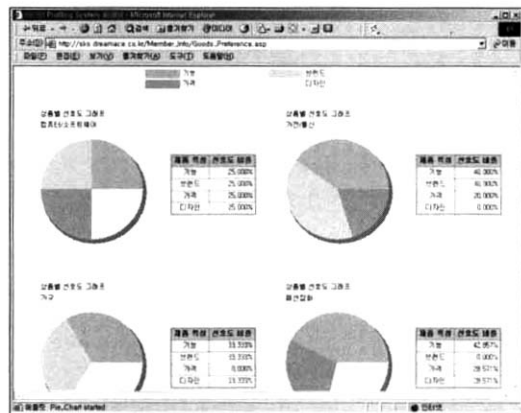
라 세부정보가 변하며, 사용자와 유사그룹의 형성 과정에 따라 갱신되어진다. 즉, 세부적인 선택과정에서의 참조되어지는 그룹에 대한 유사도 비율이 차등적으로 적용되어지는 것을 뜻한다.

사용자에 의해 선택되어진 대분류 카테고리에서의 분류는 다시 (그림 5)와 같은 유사한 사용자 그룹의 프로파일링 과정을 거쳐 유사도를 추출하게 되고 중분류 카테고리를 보이게 된다. 이러한 과정을 거쳐 최종 아이템의 선택에 이르기까지 사용자와 유사한 사용자 그룹은 축소되어지며, 상품 분류에 따른 카테고리의 특성을 이용한다할 수 있다.



(그림 6) 최종 아이টে에 대한 추천 제품과 선호도 비율

(그림 6)에서 네모 박스에 들어있는 내용이 사용자에게 제품 특성을 고려하여 추천하는 아이템이 된다. 물론 사용자는 자신이 원하는 상품을 선택할 수 있다. 사용자가 물품을 구매하면 물품의 특성을 파악하게 되고, 파악되어진 특성을 사용자가 선호하는 특성으로 간주하게 되어 사용자가 선호하는 제품군의 전체적인 특성에 반영하게 된다.



(그림 7) 사용자의 아이টে에 대한 특성 분포도

(그림 7)에서는 각 제품의 카테고리별로 사용자가 선호하는 특징들(기능성, 브랜드, 가격, 디자인)을 차트 형태로 보이고 있다. 이렇게 고려되어진 특징들은 사용자가 해당하는

품목을 선택했을 때 그 품목에서 사용자가 선호하는 특성들을 기준으로 물품을 추천해주게 된다. 또한 (그림 7)을 통해 사용자가 선호하는 특징들을 한 눈에 확인 할 수 있어, 해당 사용자의 심리를 쉽게 파악할 수 있다.

5.5 성능 평가

초기추천기능은 사용자가 처음으로 회원에 가입했을 때, 고객 성향을 파악 할 수 있는지 여부를 나타내는 기능으로 제안하는 시스템은 사용자와 유사한 신상정보를 지니고 있는 그룹을 형성하여 추천해주게 된다. 구매 전 추천은 아이템을 구매하기 이전에 아이템에 대한 사용자의 성향을 파악하는 것으로 기존 방식들은 구매를 하였을 때 또는 장바구니에 아이템을 넣었을 때에만 가능하였다. 구매 성향 파악은 사용자가 아이템을 구매했을 때, 구매한 아이템이 지니고 있는 특징을 에이전트가 파악할 수 있는지 여부를 나타내는 기능이다. 개인 성향 파악은 사용자가 아이템에 대해 지니고 있는 특징들을 파악하는 기능으로 꼭 구입하지 않고 검색만을 했어도 사용자가 구매하고자 하는 품목에 대한 특징을 분석해서 추천해 줄 수 있는 기능을 말한다. 성향성 도입 여부는 사용자가 지니고 있는 아이템 특징을 실질적으로 아이템에 대한 관심의 정도를 측정하는데 도입할 수 있는 기능을 말한다. 실시간 추천성은 웹 서핑을 모두 마쳤을 때에만 사용자의 아이템에 대한 관심 정도를 적용할 수 있는데 본 시스템은 웹 서핑 도중에도 적용할 수 있도록 설계되었다.

〈표 4〉 시스템 성능 평가

(◎: 지원 완료, △: 지원 가능, X: 지원 안 됨)

구분	제안하는 시스템	오이 뮤직	Visual Web Engine	전자 도서관 시스템
초기추천기능	△	△	X	△
구매 전 추천	◎	△	△	△
구매 성향 파악	◎	◎	△	◎
개인 성향 파악	◎	△	△	△
성향성 도입 여부	◎	△	△	X
실시간 추천성	△	X	△	X

6. 결론 및 향후 연구방향

본 논문은 사용자 개인의 관심정보와 각 제품군들에 대한 선호도 비율을 고려한 선호 상품 추천 에이전트 시스템을 제안한다.

제안하는 시스템은 사용자의 구매 상품에 대한 특성 - 구매자와 유사한 신상정보를 지닌 사용자들로부터 선호하는 상품을 추천 받을 수 있으며, 하나의 상품이라 할지라도 다양한 특성을 지니고 있다는 점 - 을 파악하여 보다 정확하게 선호하는 상품을 추천한다. 제안하는 시스템은 사용자의 구매 행위에 적용력을 가질 수 있도록 보다 정확한 사용자 프

로파일을 구축하고, 이와 같은 사용자 프로파일을 기반으로 사용자에게 불필요한 검색과정 없이 필요한 상품 정보를 제공할 수 있도록 한다.

각 고객은 본인의 개인 관심정보를 자신의 관심정보 데이터베이스에 축적하게 하는데, 이를 위해 제안하는 시스템은 웹 쇼핑을 할 때마다 관심 컴포넌트들이 재생성 되어지도록 고객 관심도의 지속적인 재계산을 통하여 관심정보를 실시간적으로 갱신한다.

본 논문은 앞에서 제시한 장점들을 구현하기 위해 시스템을 설계·제시하였으며, 사용자의 관심정보와 선호도를 계산하기 위한 알고리즘을 제시하였다. 또한 실 예를 적용하여 알고리즘의 적합성을 증명하였으며, 마지막으로 잘 알려진 '한국전산원 정보화 통계 DB 시스템'을 이용하여 제안하는 시스템을 실험하고, 대표적인 몇 가지 시스템과의 비교를 통해 성능평가를 수행하였다.

향후 연구 과제로는 각각의 아이템들에게 부여되어진 개별적인 특성까지도 고려하여 추천이 이루어진다면 사용자에게 보다 정확한 개인화된 추천을 할 수 있을 것으로 보인다. 또한 이 논문에서는 제시하고 있지 않은 항목이지만, 현재 타 연구실에서 진행되어지는 연구 중 한 부분인 "혈액형"부분도 참고를 하게 된다면 보다 효과적인 개인화된 시스템을 제시할 수 있을 것으로 사료된다.

참 고 문 헌

- [1] Trinity College Dublin, Broadcorn Eireann research Ltd., "Software Agent : A Review," May, 1997.
- [2] Maes, P. and Wexelblat, A., "Interface Agents : A Tutorial," In Conference Companion, Conference on Human Factors in Computing Systems, CHI 96, ACM Press, 1996.
- [3] Dunja Mladenic, "Personal WebWatcher : Implementation and Design," Technical report IIS-DP-7472, October, 1996.
- [4] Thorsten Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text ategorization," March, 1996.
- [5] Bruce Krulwich, "Learning document category description through the extraction of semantically significant phrases," Center for Strategic Technology Research Andersen Consulting LLP 100 South Wacker Drive, Chicago, IL 60606, 1995.
- [6] Yezdi Lashkari, Max Metral, Pattie Maes, "Collaborative Interface Agents," Conference of the American Association for Artificial Intelligence, Seattle, August, 1994.
- [7] Bruce Krulwich and Chad Burkey, "The InfoFinder Agent : Learning User Interest through Heuristic Phrase Extraction," AgentSoft Ltd., Andersen Consulting LLP, 1995.
- [8] P. Maes, R. Guttman and A. Moukas, "Agents that Buy and Sell : Transforming Commerce as We Know It," Com-

munications of the ACM, special issue on agents and electronic commerce, March, 1999.

- [9] Anthony Chavez, "Kasbah : An Agent Market-place for Buying and Selling Goods," Proceeding of the First International Conference on the Practical Application of Intelligent Agents and Multi-Agent Technology(PAAM '96), London, UK, April, 1996.
- [10] 소영준, "사용자 관심도 추출을 위한 모니터 에이전트 시스템", 석사논문, 2000.
- [11] 소영준, "사용자 프로파일 기반 개인 웹 에이전트", 정보과학회논문지, 제27권 제3호, March, 2000.
- [12] 김영설, 김병천, 윤병주, "개선된 추천시스템을 이용한 전자상거래시스템 설계 및 구현", 정보처리학회논문지D, 제9-D권 제2호, pp.329-336, 2002.
- [13] <http://www.oi.co.kr>.
- [14] 한국 전산원 정보통계 DB 시스템, "전자상거래 이용 유무에 대한 조사", <http://stat.nca.or.kr>.
- [15] 통계청, "표준 직업 분류", <http://www.nso.go.kr>.
- [16] 백혜정, 박영택, "기계학습 기반 적응형 전자상거래 에이전트 설계", 정보처리학회논문지B, 제9-B권 제6호, pp.775-782, 2002.
- [17] 성경상, "상품의 선호 특성 추출을 위한 고객 프로파일링 시스템", 석사논문, 2003.



성 경 상

e-mail : pltofgod@korea.com
 2001년 호원대학교 전자계산학과(학사)
 2003년 숭실대학교 대학원 컴퓨터학과(석사)
 2004~현재 경원대학교 대학원 컴퓨터학과 박사과정

관심분야 : 전자거래학, 유비쿼터스, 보안



박 연 출

e-mail : fearhope@empal.com
 1997년 상지대학교 응용통계학과(학사)
 1999년 숭실대학교 대학원 컴퓨터학과(석사)
 2001년 숭실대학교 대학원 컴퓨터학과 수료(박사)

1999~현재 숭실대학교 대학원 컴퓨터학과 박사과정
 관심분야 : 컴퓨터비전, 영상처리, Facial Caricaturing



안 재 명

e-mail : jmahn@sisnet.co.kr
 1988년 강원대학교 통계학과(학사)
 1995년 성균관대학교 경영대학원 정보처리 전공
 2003년 숭실대학교 대학원 컴퓨터학과 수료(박사)

2004~현재 (주)리테일테크 대표이사
 관심분야 : 유비쿼터스, 전자거래학



오 해 석

e-mail : oh@kyungwon.ac.kr
 서울대학교 응용수학과(학사)
 서울대학교 대학원 계산통계학과(석사, 박사)
 일본 동경대학 객원교수
 미국 스탠퍼드대학교 객원교수

한국정보처리학회 회장(역임)
 1982년~2003년 숭실대학교 컴퓨터학부 교수/부총장(역임)
 2003년~현재 경원대학교 소프트웨어대학 교수/부총장
 관심분야 : Multimedia, Database, 지식경영