

소프트웨어 에이전트 및 지식탐사기술 기반 지능형 인터넷 쇼핑몰 지원도구의 개발

김재경

경희대학교 경영학부 부교수
(jaek@khu.ac.kr)

조운호

동양공업전문대학 인터넷정보과 조교수
(yhcho@dongyang.ac.kr)

김우주

전북대학교 산업시스템공학과 조교수
(wjkim@chonbuk.ac.kr)

김제란

㈜오픈테크 개발본부
(jrkim@otech.co.kr)

데이터베이스 마케팅을 필두로 최근 마케팅 분야에서는 보다 고객에 적합한 제품이나 서비스를 제공하고 또한 이로 인해 그 마케팅 비용을 최소화하고 또한 그 매출 효과를 극대화하고자 하는 움직임이 가속화되고 있으며, 극단적으로는 일대일 마케팅이라고까지 표현하고 있다. 더욱이 전자쇼핑몰에 있어서는 실제 판매원이 존재하지 않는 이상 보다 더 고객의 관심을 유도하고 궁극적으로 매출을 발생시키기가 더욱 어려운 실정이며 따라서 고객을 파악하고 또한 그 고객에 적합한 제품이나 서비스에 대한 정보를 즉각적 또는 사전적으로 추측제시하여야 하는 역량이 매우 중요하다 하겠다. 그러나 이와 같은 즉시성의 추정이나 판단의 유효성을 제고하기 위해서는 전자쇼핑몰 입장에서 일단의 단편적 정보에 의존하는 방식보다는 이용가능한 모든 정보에 대한 통합적 고찰과 또한 고객에 대한 제안 여부와 추천 의사 결정을 개별적이고 순차적인 절차로 보는 관점보다는 하나의 통일된 관점에서 최대의 효과를 발생시킬 수 있도록 하는 상품 추천 방법론이 필요하다 하겠다. 본 연구는 이를 위해 전자쇼핑몰에서의 오프라인/온라인의 통합 정보를 바탕으로 추천 대상 고객 선정 및 추천 효과의 최적화를 목적으로 추천 상품 및 서비스 결정의 의사결정들에 대한 단일 의사결정 방법론 즉 상품 추천 방법론을 제안하며 이를 에이전트 기법을 바탕으로 설계하였다. 또한 이상의 방법론과 설계기법을 국내 유수의 전자쇼핑몰에 적용하여 그 실험적 성과를 제시하고 있다.

1. 서론

인터넷이나 전자상거래(E-Commerce), 콘텐츠 제공업체(Content Provider) 등이 질적으로나 수적으로 고속 성장함에 따라 쇼핑몰이나 콘텐츠 제공업체들의 경쟁이 치열해지게 되었고 이들 기업들에게는 당연히 경쟁력에 있어 상대적 우위 확보 수단에 대한 요구가 증가하게 되었다. 이에

따라 고객별로 차별화 된 마케팅을 실시하는 원투원 마케팅(One-To-One Marketing)이나 고객을 이해하고 고객과의 관계를 강화시켜 나가는 고객관계관리(CRM) 등이 기업 경쟁력의 척도로서 인식되고 있는 실정이다.

고객의 입장에서 본다면 온라인 기업이 증가함에 따라 더 많은 상품들 그리고 더 많은 공급자들이 선택가능해진 것은 사실이다. 그러나 이

* 이 논문은 2000년도 한국학술진흥재단의 지원에 의하여 연구되었음(KRF-2000-042-C00116)

와 함께 고객 역시 이들 수많은 대안들 중 어떤 대안이 정말 고객 자신의 요구에 정확히 부합되는 상품인지를 찾고 또한 평가하기 위해 더 많은 노력을 기울여야 하는 부담 역시 발생하게 되었다. 이러한 상황을 공급자 즉 온라인 쇼핑몰 입장에서 본다면 두 가지 해법이 가능하다. 첫째 보다 사용하기 쉬운 상품이나 서비스 탐색 기능을 제공하는 것이 하나이고 다른 하나는 보다 미리 고객을 파악하고 그 고객에 가장 적합한 제품이나 서비스를 추천하는 것이다. 본 연구에서는 후자의 해법에 초점을 맞추고자 하며 이와 같은 상품 추천은 다시 두가지 형태의 일련의 문제로 정의될 수 있는데 먼저 우리의 고객이 과연 누구인가 하는 문제이며 둘째는 과연 그 고객이 진정 무엇을 원하고 있는지를 파악하고 추천하는 것이다.

이와 같은 상품 추천 시스템들은 전자 상거래 환경에서 고객에게 유용한 상품을 제안하기 위해 통계적 기법과 지식탐사 기법들을 기반으로 발전하여 왔다. 기존의 오프라인(off-line) 상점에도 상품 추천 서비스가 존재하였으나 많은 경우 고객의 인구통계학적인 정보만을 기반으로 예측하는 수준이었고, 온라인(On-line) 상점의 추천 상품 서비스는 대부분 추천할 후보 상품(Candidate Product)을 고객의 구매 이력 정보에 따라 필터링하여 제공하는 방식이었다. 이러한 필터링 과정에 대한 기존 연구들은 두가지 방법으로 나누어 볼 수 있는데 내용 기반 필터링(Content-based Filtering)과 협업 필터링(Collaborative Filtering)이 그것이다.

먼저 내용 기반 필터링은 고객이 과거에 구매했던 상품들의 유사성(Similarity)을 기반으로 상품을 추천하는 것인데, 예를 들어, 강아지 사료를 구매했던 고객은 개를 키울 것이기 때문에 카펫

청소기에도 관심을 나타낼 것이라 추측할 수 있을 것이다. 한편 협업 필터링은 대상 고객과 유사한 구매 행동이나 관심을 보이는 고객들이 가장 선호하는 상품을 추천하는 방법이다. 이렇게 추천되는 상품들은 본질적인 의미에서 자신이 속한 고객 집단이 선호하는 정보이지 개인 그 자체가 선호하는 정보는 될 수 없다는 문제점이 존재한다. 또 다른 상품 추천 시스템으로 연관 규칙 알고리즘을 사용하여 과거 고객이 구매했던 상품과 연관성이 있는 다른 상품을 추천하는 방법이 있다. 특히 연관규칙 알고리즘을 통해 생성된 상품간 연관 규칙들은 개별 고객들을 위한 것이라기 보다는 상품과 상품간의 연관성만을 바탕으로 한다는 제한점이 따르고 있다.

한편 이들 기존 방법론들은 이들을 전자쇼핑몰에 적용할 경우 몇 가지 측면에서 문제점을 안고 있는데 그 첫번째가 많은 경우 고객에 대한 제한된 정보만을 이용하고 있다는 점이다. 즉 협업 필터링의 경우 오로지 고객간의 유사성 특성에만 초점을 맞추게 되며 내용 기반 필터링의 경우 오로지 구입된 상품이나 고객 자신의 특성에만 초점을 맞추게 된다. 한편 온라인 쇼핑몰에서는 대부분의 경우에 웹기반 데이터에 주로 초점을 맞추고 있다(Buchner et al., 1998; Yuan and Chang, 2001). 두번째 문제점은 대부분의 추천시스템에서 추천 자체의 정확도에만 관심을 기울이지 고객에 대한 추천 자체를 해야할 지 말아야 할지에 대한 의사 결정에는 관심을 상대적으로 기울이고 있지 않고 있다. 하지만 많은 연구자들에 의해 제기되고 있듯이 아마존과 같은 형태의 추천은 추천에 있어서의 비적시성과 과거 이력을 기반으로하는 무조건적 추천이라는 문제점을 드러내고 있다. 실제로 추천 자체가 해당 고객에게 적합하지 않거나 시기적절하지 않을 경우 오히려

전체적 쇼핑물의 이미지에 역효과를 낼 수 있으며 따라서 특정 고객별로 추천을 할 것인지에 대한 본질적 고찰이 우선되어야 할 것이다. 마지막으로 이러한 일련의 추천 관련 의사 결정들이 통합적 틀 안에서 연속적인 추천 업무를 수행할 수 있도록 이루어져야 하며 동시에 다양한 추천의 채널을 수용할 수 있는 시스템적 설계 측면도 고려되어야 할 것이다.

따라서 본 연구에서는 고객이 전자쇼핑물에서 상품을 처음 인식하는 행동에서부터 상품을 클릭하고 장바구니에 담아 상품을 구매하기까지의 과정에 대한 정보와 같은 웹기반 정보뿐만 아니라 고객의 지리적, 환경적, 및 개인적 특성에 관련된 오프라인 정보들을 통합적으로 고려하고자 하며 또한 고객이 원하지 않는 상품을 추천하게 될 경우, 고객의 신뢰도를 떨어트릴 수 있기 때문에 이러한 오류를 방지하기 위해 예측 모델링 기법을 사용하여 추천된 상품을 구매할 가능성이 있는 고객만을 선별하고자 한다. 특히 선별된 대상 고객들에게는 연관 규칙 알고리즘을 통해 얻어진 상품 클래스간의 연관성을 바탕으로 해당 고객이 가장 선호할만한 상품 클래스를 할당하며, 선택된 상품 클래스에 속하는 상품들 중에서 구매 효율성이 가장 높은 상품이 추천함으로써 상품의 추천 효과를 극대화하고자 한다. 또한 이와 같은 방법론을 보다 현실적으로 구현할 수 있도록 에이전트를 기반으로 하는 상품 추천 대상 결정 및 추천 시스템의 통합 시스템 설계를 수행하였다. 특히 에이전트 구조를 채택함으로써 데이터 마이닝 전문가와 마케터의 역할을 명확히 구분하였으며 따라서 데이터 마이닝을 자세히 알지 못하는 마케터일지라도 데이터 마이닝 전문가가 구성한 정보를 이용하여 쉽게 사용할 수 있도록 구성되어 있다.

이제 다음의 2장에서는 상품 추천 시스템에 관련된 기존의 연구들을 간략히 살펴보고, 본 연구의 상품 추천 시스템의 이해를 돕기 위해 상품 추천 시스템에 활용된 연관 규칙 탐사 기법과 예측 모델링 기법들과 그 외 웹로그를 활용한 연구에 대해서 소개하고자 한다. 이어 3장에서는 본 연구에서 제안하고자 하는 상품 추천 시스템의 방법론을 예제와 함께 기술하고자 하며, 4장에서는 에이전트를 기반으로 하는 상품 추천 통합 시스템 구조를 소개하고 그 설계 이슈들을 다루고자 한다. 그리고 5 장에서는 제안한 방법론과 그 구현 시스템을 이용한 실제 운영중인 국내 유수의 쇼핑몰 사이트 적용 사례를 소개하고자 하며 마지막으로 5장에서는 현재까지의 연구 결과로부터의 결론 및 향후 연구 방향을 제시하고자 한다.

2. 관련 연구

최근 들어 웹을 기반으로 하는 상품 추천 시스템에 대한 많은 연구가 이루어지고 있는데 (Konstan et al., 1997; Wolf et al., 1999; Borchers et al., 1998; Resnick and Varian, 1997), 대부분의 이러한 웹기반 상품 추천 시스템들은 통계적 기법이나 지식 탐사 기법을 적용하여 고객이 선호하는 상품에 대한 정보를 파악한 후 해당 고객들에게 이메일 등이나 온사이트 추천이나 배너등의 채널을 통해 적합한 상품을 추천하고 있다. 본 장에서는 고객에게 유용한 추천을 해주기 위한 시스템들의 방법론으로 많이 이용되고 있는 협업 필터링 상품 추천 기법과 연관 규칙 기반 상품 추천, 그리고 이 두 가지를 결합한 추천 기법들에 대해 살펴보고자 한다.

이외에도 본 연구에서 상품 추천 과정에 통합적으로 고려하고자 하는 상품 추천 대상 고객 선정과 관련된 기존 예측 모델링 기법들에 대해 알아보하고자 하며, 마지막으로는 웹로그 분석과 이의 활용 연구들을 간략히 소개하고자 한다.

2.1 협업 필터링 상품 추천 시스템

먼저 협업 필터링(Collaborative filtering) 방법론(Resnick et al., 1994; Konstan et al., 1997; Sarwar et al., 2000)은 웹을 기반으로 하는 전자쇼핑몰에서 이용되고 있는 성공적인 상품 추천 시스템 기법중의 하나라 할 수 있으며, 대상 고객과 유사한 구매 이력을 보이는 고객들의 상품에 대한 선호를 바탕으로 대상 고객에게 유용한 상품을 추천하는 방법으로써, 대상 고객에 대한 유사 고객들을 찾기 위해 통계적인 기법 등을 중심으로 여러 다양한 기법들을 응용하고 있다. 또한 이렇게 대상 고객의 유사 고객 집단이 결정된 후 다시 여러가지 알고리즘들을 이용하여 해당 고객에 대한 추천 상품을 결정하게 된다. 일반적으로 이러한 협업 필터링 상품 추천 과정은 크게 입력 데이터 구성, 이웃 집단 탐색, 추천 상품 결정 단계로 나뉘볼 수 있는데 이들 각각은 일반적으로 다음과 같이 구성된다(Sarwar et al., 2000).

(1) **입력 데이터 구성(Representation)** : 협업 필터링을 기반으로 하는 상품 추천 시스템에서의 입력 데이터는 보통 n 개의 상품에 대하여 m 명의 고객의 구매 트랜잭션의 집합으로 구성되며 따라서 이는 보통 $n \times m$ 의 임의의 고객-상품 행렬 R 로 표현할 수 있다. 이 행렬의 개별 i 번째 행 j 번째 열의 값 r_{ij} 는 i 번째 고객이 j 번째 상품을 구매했을 경우는 1의 값을 그렇지 않으면 0의 값을

각각 가지게 된다.

(2) **유사 집단 탐색(Neighborhood Formation)** : 고객간의 유사성을 계산하는 유사 집단 탐색 단계는 협업 필터링 기반 상품 추천 방법론에서 가장 중요한 단계라 할 수 있는데, 유사 집단을 탐색하는 주요 목적은 정형적으로 정의하지 않으며 각 고객 u 에 대하여 가장 근접한 1명의 고객 리스트 $M = \{M_1, M_2, \dots, M_l\}$ 을 찾는 것이라 할 수 있다. 여기서 만약 $u \notin M$ 이고 $sim(a, b)$ 가 고객 a 와 고객 b 의 유사성 정도를 나타내는 함수라면, $sim(u, M_1) > sim(u, M_2) > \dots > sim(u, M_l)$ 임을 이 집합 M 은 의미하게 된다. 이러한 유사 집단 탐색 단계는 먼저 고객간의 유사도를 측정 한 이를 바탕으로 유사 집단들을 결정하게 되는데 대부분의 경우 두 고객 a 와 b 의 근접도를 측정하는 방법으로써 피어슨 상관관계수(Pearson Correlation), 코사인 프로젝트션(Cosine Projection) 등을 사용하고 있다. 이렇게 모든 고객간의 유사도를 계산한 후 유사 집단을 결정하여야 하는데 간단하게는 대상 고객과 가장 가까운 1명의 고객들을 선택하여 결정하는 방법이 있다.

(3) **추천 상품 결정(Generation of Recommendation)** : 상품 추천을 위한 마지막 단계로서 설정된 유사 집단으로부터 상위 N 개의 추천 상품을 이끌어 내는 단계이며, 추천 상품들을 판단하기 위한 기준으로서 보통 빈발 구매 상품 추천(Most-frequent Item Recommendation)과 연관 규칙 기반 추천(Association Rule-based Recommendation) 방법 등이 있다. 빈발 구매 상품 추천은 대상 고객의 유사 집단에 대한 구매 이력 데이터를 분석하여 상품별 구매 빈도 수가 높은 상위 N 개의 상품을 추천하는 방법이고, 연관 규

칙 기반 추천은 대상 고객의 유사 집단을 대상으로 하여 생성된 상품 연관성 규칙을 적용하여 연관성이 높은 상위 N 개의 추천 상품을 제시하는 방법이다.

2.2 연관 규칙을 이용한 상품 추천 시스템

데이터로부터 발견적 지식을 추출하는데 주요 목적을 갖는 지식탐사 기법(KDD: Knowledge Discovery and Data Mining) 역시 고객에게 상품을 추천하기 위한 지식을 추출하고 이용하기 위해 많이 이용되고 있으며 이들 기법 중 가장 상품 추천에 관련 깊은 기법 중 하나는 바로 연관 규칙 탐사 기법(Association Rule Mining)이다. 이제 아래에서 어떻게 연관 규칙 탐사 기법이 상품 추천에 이용되는지에 대해 간략히 알아보고자 한다.

(1) 연관 규칙 탐사 기법(Association Rule Mining) : 연관 규칙 탐사 기법은 데이터베이스에 존재하는 여러 항목들간 또는 개별 값들간의 확률적 연관성을 탐사하여 임의의 유의성 이상을 갖는 패턴들을 결과적으로 추출하는 기법으로서 1993년에 처음 소개되었고 특히 장바구니 분석으로도 많이 알려져 있다(Agrawal et al., 1993; Agrawal and Srikant, 1994). 먼저 n 개의 상품들로 구성된 집합 $P = \{P_1, P_2, \dots, P_n\}$ 가 존재한다고 가정하자. 또한 임의의 상품 집합 즉 트랜잭션 $T \subseteq P$ 는 어떤 고객이 한번에 구매한 상품들의 집합으로 정의하고자 한다. 이 때, $X, Y \subseteq P$ 와 $X \cap Y = \emptyset$ 를 만족하는 임의의 상품 집합 X 와 Y 간의 연관 규칙이란 미래의 발생가능한 임의의 상품 집합 T 에서 집합 X 의 상품들이 존재한다면 곧 T 에 집합 Y 의 상품들이 존재할 가능성

이 높다는 것을 의미한다. 일반적으로 이러한 연관 규칙을 우리는 $X \Rightarrow Y$ 의 형식으로 표현하고 있다.

이와 같은 연관 규칙들에 대한 연관성의 정도는 보통 지지도(Support)와 신뢰도(Confidence)를 바탕으로 평가하고 있는데, 지지도 s 란 해당 연관 규칙 패턴의 발생 빈도를 나타내고, 신뢰도 c 는 규칙의 확률적 강도를 의미한다. 이들을 좀 더 정형적으로 정의한다면 먼저 임의의 연관 규칙 $X \Rightarrow Y$ 에 대한 지지도는 다음의 공식 (1)과 같이 전체 트랜잭션 중에서 X 와 Y 를 포함하는 트랜잭션의 비율로 측정한다.

$$s = \frac{\text{number of transactions containing } X \cup Y}{\text{number of transactions}} \quad (1)$$

한편 신뢰도 c 는 다음의 (2)와 같이 정의할 수 있으며 이는 X 를 포함하는 모든 트랜잭션 중 $c\%$ 만이 Y 를 포함하고 있다는 것을 의미한다.

$$c = \frac{\text{number of transactions containing } X \cup Y}{\text{number of transactions containing } X} \quad (2)$$

일반적으로 신뢰도가 높은 규칙은 다른 규칙에 비해 예측력은 높다고 할 수 있지만, 발생 빈도가 낮을 경우 해당 규칙의 유용성이 떨어지므로 보통 연관 규칙 추출에 있어 이용자가 제시하는 최소 지지도와 최소 신뢰도를 모두 만족하는 연관 규칙만을 대상으로 하게 된다.

(2) 상품 클래스 단위의 연관 규칙 : 이상에서 다룬 연관 규칙을 개별 상품들에 적용할 경우 좀 더 구체적 정보를 이끌어 낼 수 있을지 몰라도 실제로 유의한 규칙들을 발견하기가 어려운 것은

사실이며, 따라서 개별 상품간의 연관 규칙보다는 상품 클래스간의 연관 규칙을 추출하는 것이 오히려 더 실질적인 의미를 갖는 경우가 많다 (Agrawal et al., 1993; Resnick et al., 1994). 그러나 한편으로는 상품의 클래스가 너무 추상적일 즉 상위의 개념일 경우, 결과적으로 생성되는 규칙들이 이전 너무도 당연하고 상식적인 수준이어서 적용하기에 부적합할 수도 있는 것이다 (Klemettinen et al., 1994). 이와 함께 최근 연관 규칙의 조건부와 결과부의 추상적 수준이 서로 다른 다단계 연관 규칙을 추출하기 위한 여러 추가적 알고리즘도 연구되어오고 있으며(Han and Fu, 1999; Adomavicius and Tuzhilin, 1999), 이러한 다단계 연관 규칙은 동일 수준의 개별 상품이나 클래스간의 연관 규칙으로 제한되어 있지 않고 서로 다른 다른 수준의 상품 클래스 또는 상품들간의 연관 규칙에 대한 탐색을 허용함으로써 상대적으로 흥미롭고 유용한 연관 규칙들을 발견할 가능성이 높다는 장점을 가지게 된다.

(3) **고객별 상위 N개의 추천 상품 생성** : 먼저 연관 규칙은 상품 추천 방법론 관점에서 볼 때, 임의의 고객 u 에 대한 추천 효과 면에서 상위 N 개의 추천 상품들을 구성하는데 이용될 수 있다 (Sarwar et al., 2000). 이를 위해 우선 고객 u 에 관련하여 미리 결정된 지지도 이상을 나타내는 연관 규칙을 탐사하고, 탐사된 연관 규칙의 조건부에 해당하는 상품들을 고객이 모두 구매한 경우에 한하여, 결론부에 나타나고 동시에 아직 고객 u 에 의해 구매가 이루어지지 않은 상품이나 상품 클래스를 추출하게 된다. 이렇게 추출된 상품들을 다시 이들 상품들을 추출하기 위해 이용된 연관 규칙들의 신뢰도를 기준으로 정렬하고 이중 신뢰도 관점에서 상위 N 개의 상품을 추천

상품으로 선택하게 되는 것이다. 이 때, 만약 특정 상품이 여러 연관 규칙들에 의해 추출되었다면, 그 중 가장 높은 신뢰도를 갖는 연관 규칙의 신뢰도를 해당 상품에 적용하게 된다.

이상에서 살펴본 바와 같이 협업 필터링이나 연관 규칙만을 기반으로 하는 상품 추천 방법은 전체 구매 고객 혹은 유사 구매 고객 집단의 구매 데이터만을 이용할 뿐, 개별 고객의 상품 선호도를 추천 과정에 반영하지 못하는 단점을 가지고 있다. 한편 이와 같은 단점을 해결하기 위한 연구들로 이루어져왔는데 다음에서 한 대표적 관련 연구를 소개하고자 한다.

2.3 협업 필터링과 연관 규칙을 통합한 상품 추천 시스템

Lawrence와 그 동료들은 개별 구매자의 상품 선호도를 보다 적극적으로 상품 추천에 반영하기 위하여 협업 필터링 기법과 연관 규칙 기법을 결합한 오프라인(Off-Line) 상점을 위한 상품 추천 시스템을 제안하였다(Lawrence et al., 2000). 이 연구는 먼저 유사한 구매 이력을 갖는 고객들을 동일 집단으로 구분하고, 특정 고객에게 그 고객이 속한 집단의 인기 상품들 중에서 해당 고객의 선호도를 기반으로 흥미를 줄 수 있는 상품들을 추천하는 방법을 제안하였는데, 이 때, 협업 필터링은 유사 고객 집단의 탐사에, 연관 규칙 기법은 개별 상품 클래스간의 연관성 탐사에 사용되었다.

각 고객의 상품 선호도는 그들의 이전 구매 이력 정보를 바탕으로 아래와 같은 고객의 구매 벡터(Spending Vector)로서 표현되며, 이 때, 고객 m 에 대한 구매 벡터(Spending Vector) $C(m)$ 은 다음과 같이 정의할 수 있다.

$$C^{(m)} = [C_{m1}, C_{m2}, \dots, C_{ms}, \dots, C_{mS}]^T, m = 1, \dots, M$$

여기서, C_{ms} 는 상품 클래스 s 에 속하는 상품들에 대한 구매 회수를 의미하고, M 과 S 는 각각 총 고객 수와 상품 클래스의 수를 의미한다. 이 구매 벡터를 구체적으로 이용하기 위해 다시 아래의 식과 같은 정규화(Normalization)과정을 거쳐 유사 고객군 탐사 알고리즘에 이용되게 된다.

$$\hat{C}_{ms} = \frac{C_{ms}}{\sum_{s'=1}^S C_{ms'}}$$

유사 고객 집단의 분류가 완료되면, 각 고객 집단 별로 소속 고객들이 자주 구매하는 상품 집합을 구성하게 되며, 이렇게 얻어진 각 고객 집단별 인기 상품 집합들은 특정 대상 고객의 상품 선호도와 연계하여 추천 상품들의 우선 순위를 결정하는 작업을 수행하는 매칭 엔진(Matching Engine)에 전달되게 되는 것이다. 이 때, 매칭 알고리즘(Matching Algorithm)은 고객의 구매 벡터와 각 관련 상품 벡터(Product Vector)들간의 유사(Similarity) 정도에 따라 상품들에 해당 추천 가능 점수를 할당하게 된다. 특히 매칭 알고리즘에서 연관 규칙 탐사 기법의 역할은 각 상품의 벡터를 구성함에 있어 상품 클래스간 연관성도 고려하고 있으므로 해서, 이들 상품 클래스간의 연관성 정보를 탐색하는데에 있다.

먼저 각 상품 $n = 1, \dots, N$ 에 대한 상품 벡터 $P^{(n)}$ 를 다음과 같이 정의할 수 있다.

$$P^{(n)} = [P_1^{(n)}, P_2^{(n)}, \dots, P_s^{(n)}, \dots, P_S^{(n)}]^T$$

이 때, $S(n)$ 을 상품 n 이 속한 상품 클래스라 하

고, $C(s)$ 를 상품 클래스 s 가 속한 상위 상품 클래스라 한다면, 위 상품 벡터를 구성하는 각 $P_s^{(n)}$ 들은 역시 다음과 같이 정의할 수 있다.

$$P_s^{(n)} = \begin{cases} 1.0 & : \text{if } s = S(n) \\ 1.0 & : \text{if } S(n) \Rightarrow s \\ 0.5 & : \text{if } C(s) = C(S(n)) \\ 0.25 & : \text{if } C(S(n)) \Rightarrow C(s) \\ 0. & : \text{otherwise} \end{cases}$$

다시 말해서, $P_s^{(n)}$ 는 개별 상품 n 이 특정 상품 클래스 s 와의 연관성이 어느 정도인지를 나타내고 있으며, 임의의 고객 m 과 상품 n 의 연관성을 나타내는 유사 점수(Score), σ_{mn} 는 두 벡터, $C^{(m)}$ 와 $P^{(n)}$ 의 코사인 계수(Cosine Coefficient)를 구하여 계산하게 된다.

$$\sigma_{mn} = \rho_n \frac{C^{(m)} \cdot P^{(n)}}{\|C^{(m)}\| \|P^{(n)}\|}, \text{ where } \rho_n = \left[\frac{PM_n}{\overline{PM}} \right]^\alpha$$

여기서 ρ_n 은 조절 인자(Modulation Factor)로서 각 상품의 재고 유지 비용이나 이윤 등과 같은 마케팅 조건을 반영하기 위해 사용되며, 상품 벡터는 상품 클래스 단위로 생성되기 때문에 각 상품 $n = 1, \dots, N$ 에 대해서 $\rho_n = 1$ 이라면 특정 고객에 대해서 동일한 상품 클래스에 속하는 모든 상품들은 같은 점수가 부여되게 되는 것이다. 따라서 ρ_n 은 동일한 상품 클래스에 소속된 각 상품들을 차별화하기 위한 요소로서 ρ_n 의 계산에 이용되는 PM_n 은 상품 n 과 연관된 이윤(Profit Margin)을 의미하고, \overline{PM} 은 전체 상품으로부터의 평균 이윤을 의미한다. 또한 α 는 전체적으로 조절 인자의 영향력 정도를 제어하기 위한 경험

적인 인자(Empirical Factor)로 활용되고 있다.

2.4 웹로그 활용에 관한 연구

인터넷 쇼핑물에 있어서 획득가능한 고객 정보에는 오프라인 기업에서와 같은 구매 이력외에도 고객들이 쇼핑 물 사이트를 방문할 때 발생하는 웹로그 정보가 있으며, 이는 인터넷 기업에 있어 고객 분석시 간과해서는 안될 매우 중요한 정보인 것이다. 특히 상품 추천문제의 경우, 고객의 성향이나 상품 선호도 분석에 있어 핵심적인 기초 정보로 활용하여야 하며 따라서 본 연구에서는 이러한 웹로그 정보를 적극적으로 상품 추천에 활용하고자 하며 본 절에서는 간략히 관련된 개념들을 소개하고자 한다.

(1) 클릭스트림(Clickstream)

클릭스트림은 웹사이트에서 방문객의 이동 경로를 말한다. 인터넷 쇼핑물 관점에서 이러한 클릭스트림은 고객이 어떻게 온라인 상점을 찾아오게 되는지, 어떤 상품을 살펴보았으며, 결과적으로 어떤 상품을 구매하였는지 등 고객 행태 정보를 제공하기 때문에 클릭스트림에 내재된 정보를 분석하는 것은 인터넷 쇼핑물에서 매우 중요하다 하겠다. 전통적으로는 사이트의 마케팅 수행 성능을 측정하기 위하여 방문 고객 중 상품을 구매한 고객들의 비율만을 이용하였으나, 이 비율은 전체 쇼핑물의 효과를 평가하는 데 유용하지만, 쇼핑물 내에서의 개별 프로세스에 대한 성과를 측정하는데는 미흡하였다. 따라서 이를 좀 더 세분화한 클릭스트림들에 대한 정의가 제안되기도 하였다(Lee and Podlaseck, 2000).

Look-to-Click Rate: 상품 배너에 노출된 경우

이를 클릭하게 되는 비율

Click-to-Basket Rate: 상품 배너에 연결된 상품 소개 정보를 본 경우 이를 장바구니에 넣게 되는 비율

Basket-to-Buy Rate: 장바구니에 넣은 상품들에 대하여 구매하게 되는 비율

Look-to-Buy Rate: 위의 세 과정을 전체적으로 평가하는 비율로서 상품 배너에 노출된 경우 결과적으로 이를 구매하게 되는 비율

이와 같은 세분화된 클릭스트림들은 개별 배너의 효과, 상품 정보의 유효성, 장바구니 상에서의 행태 등을 파악할 수 있게 해주며, 따라서 보다 분석적으로 인터넷 쇼핑물의 업무 분석 및 고객 행태 분석에 이용될 수 있게 해주고 있다.

(2) 웹로그 분석(Web Log Analysis)

이외에도 일반적인 웹로그 분석이라 함은 웹로그 데이터를 이용하여 웹 사이트 방문자들의 다양한 사이트 방문 행태를 파악하고 웹사이트의 페이지뷰, 사용자별 페이지뷰, 접속장소 및 방식, 시간별 페이지뷰, 방문자수 등에 대한 현황 및 추세를 분석할 수 있도록 하는 방법론을 말하며 (Cooley et al., 1999), 광의로 해석하자면 단지 웹로그 데이터뿐만 아니라, 웹사이트에서 보유하고 있는 고객등록정보, 구매정보, 외부 정보 등과 함께 통합된 형태의 데이터를 복합적으로 사용하는 분석을 웹로그 분석이라 할 수도 있겠다. 본 연구에서는 보다 효과적인 고객에 대한 상품 추천을 위하여 광의의 웹로그 분석을 하부 과정으로 포함시키고 있으며 이제 다음 장부터는 이러한 통합 데이터에 근거한 상품 추천 방법론을 소개하고자 한다.

3. 추천 대상 고객 선정 및 상품 추천 방법론

3.1 상품 추천 문제

상품 추천방법론을 제안하기에 앞서 먼저 여러가지 형태의 상품 추천 상황이 존재할 수 있으며 이에 따라 여러가지 상품 추천 방법이 각 상황에 따라 서로 다르게 나타날 수 있음을 밝히고자 한다. 이러한 상품 추천 상황은 첫째, 대상 고객이 전 고객을 상대로 할 것인지 아니면 특정 선택된 고객을 대상으로 할 것인지에 따라 구분될 수 있으며 다시 둘째로 추천 대상 상품의 수에 제한이 있는 지에 따라 구분할 수 있다. 마지막으로 추천 상품의 추천 시기가 결정된 경우인지 아니면 상시 추천인지에 따라 역시 그 상품 추천 상황을 나누어 볼 수 있을 것이다.

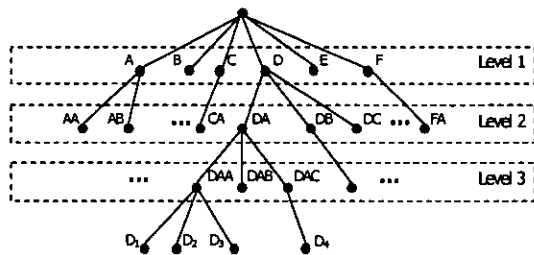
본 연구에서는 이상의 가능한 추천 상황 중 선택적 고객에 대한 유한한 수의 상품 추천을 기획된 시기에 수행하고자 하는 상황에 대한 상품 추천 문제를 다루고자 한다. 선택적 고객의 의미는 잘못된 추천으로 인한 음의 효과가 있음을 말하는 것이고 유한한 수의 상품 추천이란 추천 가능 공간에 제약이 있는 경우로서 일반적으로 유형의 카탈로그나 적절한 규모를 유지해야하는 이메일에 의한 추천 등이 이에 해당될 것이다. 마지막으로 기획된 시기의 추천이란 캠페인적 상품 추천 활동을 의미한다고 볼 수 있다. 이러한 상품 추천 상황외에도 상품에 대한 정의 또는 추천의 수준에 의하여 추천 문제를 구분할 수도 있는데 일반적으로 쇼핑몰에서의 상품은 <그림 1>에 나타난 바와 같이 개별 단품으로부터 출발하여 계층적 상품 분류 체계를 이용하여 분류할 수 있다. 따라서 상품 추천의 수준이나 추천을 위한 분석

과정에서 어떻게 이러한 분류 체계를 사용하는가에 따라 추천 문제를 구분할 수도 있는 것이다.

이미 많은 기존 연구들에서도 지적되었듯이 각 단품간의 연관성을 바탕으로 상품 추천을 수행할 경우 유의한 연관성을 획득하기가 매우 어려우며 또한 상위의 계층을 바탕으로 연관성을 탐색할 경우 너무 당연한 연관성만을 획득하기가 쉬우므로 본 연구에서는 이러한 분석 대상 상품 계층의 결정을 분석자의 의지에 맡기는 전략을 취하고자 한다.

이상에서의 논의를 바탕으로 본 연구의 대상 상품 추천 문제를 정형적으로 표현한다면 Recommendation(p, t, n, l)으로서 나타낼 수 있는데, 여기서 p 는 최소 p 개 이상의 상품 클래스에서의 상품들을 구매한 고객을 대상으로 한다는 것을 의미하며, t 는 t 시점에서의 상품 추천을 위한 추천 문제라는 것을 의미하며, n 은 n 개의 상품을 추천하고자 한다는 것이며 마지막으로 l 은 상품 계층도상 수준 l 에 대한 상품 연관성 분석을 바탕으로 할 것이란 것을 의미한다.

예를 든다면 Recommendation(2, 2001/12, 2, 3) 문제는 <그림 1>에서의 3수준에 해당하는 상품 클래스에서 최소 2개 이상의 상품 클래스에 대한 구매를 기록한 고객에 대해 2001년 12월에 각 고객별로 2개의 제품을 추천하고자 하는 문제로서 해석할 수 있는 것이다.



<그림 1> 상품 분류 체계

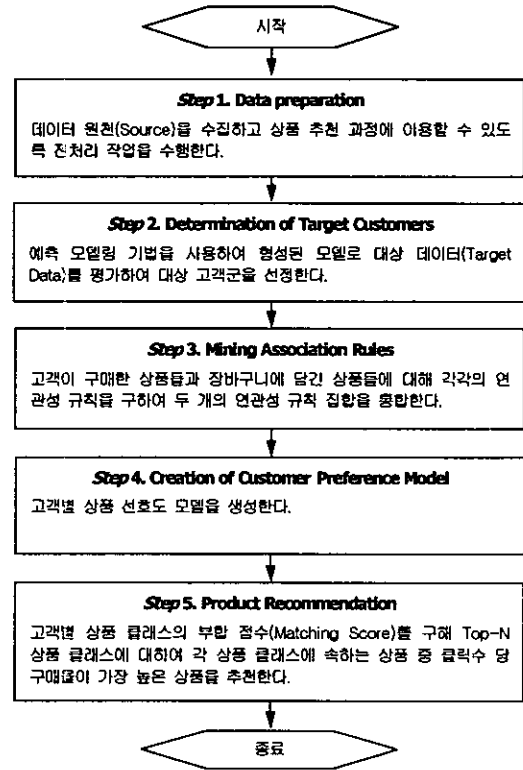
3.2 상품 추천 방법론

본 연구에서 제안하고자 하는 상품 추천 방법론은 앞에서 정의한 상품 추천 문제를 해결하기 위한 방법론으로서 기존의 상품 추천 방법론들에서 발견되는 여러가지 단점들을 보완하기 위한 다음과 같은 세가지 세부 목적들을 달성하고자 하였다.

- (1) 무작위적 상품 추천으로부터 발생하는 부정적 효과를 최소화한다.
- (2) 고객의 상품 선호도를 반영하여 추천 효과를 극대화한다.
- (3) 상품 추천에 의한 쇼핑몰의 수익성을 최적화한다.

첫번째 목적을 달성하기 위하여 본 연구에서는 일반적인 상품 추천 방법론과는 달리 상품 추천 고객의 선정 방법론을 상품 추천의 한 하부과정으로 포함시켰으며 이를 위한 절차와 방법론을 따로 제안하였다. 두번째 목적을 달성하기 위해서는 Lawrence (2000)의 방법론과 유사한 고객의 상품 선호도 측정 방법을 개발하였으며 이에 따른 추천 상품 평가 방법을 제안하였다. 또한 세번째 목적을 위하여 상품의 구매 효율성을 극대화시키는 방향으로 최종 추천 상품을 결정하는 방법론의 제안하고 있다.

따라서 제안하고자 하는 상품 추천 방법론의 전체적 구성은 <그림 2>에서 볼 수 있는 바와 같이 다섯 단계의 순차적 과정으로 구성되며 이들은 각각 자료 준비 단계, 추천 대상 고객 선정 단계, 상품 클래스간 연관성 분석 단계, 고객 상품 선호도 분석 단계, 상품 추천 단계이다. 이제 이들 각 단계에서의 연구 이슈와 구체적 방법론에 대하여 소개하고자 한다.



<그림 2> 상품 추천 방법론의 각 단계들

3.3 자료 준비 단계

인터넷 쇼핑몰에서의 상품 추천 과정을 수행하기 위해서는 무엇보다도 먼저 상품 추천 과정에 필요한 데이터가 수집되어야 한다. 본 연구에서는 우선 인터넷 쇼핑몰에서 일반적으로 획득가능한 고객에 대한 데이터 원천으로서 가장 기본적인 고객 데이터, 상품 데이터, 판매 데이터와 웹서버에 저장되어 있는 웹로그 데이터를 이용하고자 한다.

(1) 고객 데이터(Customer Data)

고객 데이터란 고객의 나이, 성별, 학력, 결혼

여부, 직업 그리고 소득과 같은 인구통계학적 정보와 사회계층, 라이프 스타일, 개성과 같은 심리 분석적 정보들을 의미하는데 이러한 정보들은 일반적으로 인터넷 쇼핑물에서의 고객 등록 정보에 의해 획득가능한 정보를 의미하기도 한다.

(2) 상품 데이터(Product Data)

상품 데이터란 상품명, 상품 가격, 브랜드, 제조사 및 상품 속성등에 대한 일반적인 상품 정보를 의미하며 이 역시 일반적으로 인터넷 쇼핑물에 있어서 내부 데이터베이스에서 획득가능한 정보이다. 궁극적으로 이러한 상품 데이터는 뒤에서 소개할 판매 데이터와 결합하여 상품별 판매 횟수, 브랜드 선호도 등의 이차적 정보를 추출하는데 사용되게 된다.

(3) 판매 데이터(Sales Data)

판매 데이터는 일반적으로 고객 식별자, 판매 트랜잭션 식별자, 구매 날짜 및 구매 상품 등의 속성들로 구성되며 역시 데이터베이스를 이용하여 저장 운영된다. 특히 판매 데이터로부터 추출 가능한 고객별 구매 이력 정보는 고객의 상품 구매 행태나 고객의 상품 선호도를 판단하는데 있어 중요한 요인으로서 이용될 수 있다.

(4) 웹로그 데이터(WebLog Data)

고객이 인터넷 쇼핑물의 웹 사이트를 방문하게 되면 이 때, 해당 웹 서버에는 자동적으로 고객의 방문에 관련된 로그가 저장되게 되는데, 이러한 웹로그 데이터의 분석은 앞에서 소개한 다른 유형의 데이터와 함께 고객의 상품 선호도를 파악하는데 있어 매우 중요한 역할을 수행하게 된다. 그러나 이러한 웹로그 데이터는 기본적으로

로 앞에서의 세가지 유형의 데이터들과는 다르게 데이터베이스에 의하여 관리된다기보다는 단순 텍스트 파일 형식으로 유지됨으로써 이를 추가적으로 파싱하고 분석하여야 하는 작업이 필요하다. 본 연구에서 역시 이러한 웹로그 전처리 기법을 이용하여 앞의 세가지 데이터와 함께 다음에서 소개할 상품 추천 데이터마트(Data Mart)를 구축하고 있다.

본 연구에서는 이상에서 소개한 네가지 유형의 정보들을 바탕으로 상품 추천을 위해 사용할 데이터들을 전처리 작업(Berson and Smith, 1997)을 통하여 데이터마트를 구축하였으며 크게 고객 속성 정보, 거래 및 장바구니 정보, 클릭스트림 정보 등으로 나누어지며 각각의 구성은 대략 다음과 같다.

(1) 고객 속성 정보

먼저 고객 속성 정보는 <표 1>에서와 같이 다시 인구통계학적 정보, 행동학적 정보, 심리학적 정보 등으로 나누어지는데 인구통계학적 정보와 심리학적 정보는 고객 등록 시점에서의 등록 화면과 설문 화면을 통하여 습득되고 유지되며 따라서 이는 일반적으로 데이터베이스로부터 단순 획득 가능하며, 행동학적 정보는 고객이 상품을 구매하는 행태 및 웹상에서 행동 패턴 등에 대한 정보를 말하며 구체적인 예로서는 최근 주문일, 최근 방문일, 구매 회수 등을 들 수 있으며 이는 판매 데이터와 웹로그 데이터를 분석하여 획득하였다.

(2) 거래(Transaction) 및 장바구니 정보

거래 정보는 판매 데이터(Sales Data)에서 연 관 규칙 탐사 기법을 적용하는데 필요한 데이터만을 선택 추출하고 이를 적절한 형태로 변형한

후, 이를 위해 고안된 특수한 트랜잭션 식별 알고리즘을 이용하여 구성된다. 장바구니 정보 역시 유사한 웹로그 분석 과정을 통해 추출되어지는 정보로써 실제 구매가 이루어졌는지 여부에 관계 없이 고객의 장바구니에 대한 추적 정보로서 구성된다.

(3) 클릭스트림 정보

클릭스트림 정보는 거래나 장바구니 이외의 고객의 행태를 의미하는 정보로서 고객이 웹사이트를 방문하여 상품을 구매하기까지 보인 행동을 추적 측정하게 되는데, 본 연구에서는 이를 위해 인터넷 쇼핑물에 가장 일반적인 ‘상품 확대보기’, ‘장바구니에 담기’, ‘상품 주문하기’ 등의 세가지 행태로 구분하고 이에 관련된 추적 정보를 역시 웹로그를 분석하여 추출하는 알고리즘을 개발 적용하여 해당 데이터를 추출하였다.

<표 1> 고객 속성 정보 분류

속성 유형	데이터 원천	Examples
인구 통계학적	고객 데이터	성별, 나이, 직업, 수입, 지역, 교육정도, ...
행동학적	판매 데이터	최근 주문일, 총구매 금액, 구매회수, 취소건수, ...
	웹로그 데이터	최근 방문일, 방문수, 머문시간, 장바구니에 담은 상품 수, ...
심리학적	설문지 조사	개성, 라이프 스타일, ...

3.4 추천 대상 고객 선정 단계

상품 추천을 위한 자료가 준비된 다음 본 연구의 상품 추천 방법론은 추천 대상 고객 선정 단계를 수행하게 된다. 앞에서 언급한 바와 같이 본 연구는 무작위적 추천으로 인한 부정적 효과

를 최소화하기 위해 추천에 의한 구매 가능성이 높은 고객을 먼저 선정하고 이들에 대한 선택적 추천을 수행하고자 한다. 이를 위해서는 먼저 추천 대상 고객 선정 모델이 필요하며 이러한 모델을 구축하기 위한 절차 역시 수립되어야 한다.

추천 대상 고객 선정 모델로서는 의사결정나무(Decision Tree) 방법론(Quinlan, 1986; Quinlan, 1993)을 채택하였으며 이 모델을 이용하여 각 고객의 해당 시점 t 에서의 구매 가능성을 구매 가능 고객과 구매 불가능 고객으로 분류함으로써 추정코자 하였다. 이와 같은 추천 대상 고객 선정 모델을 구축하기 위하여 우리는 먼저 모델 세트(Model Set)와 스코어 세트(Score Set)라는 고객 데이터 집합들을 상품 추천 문제 Recommendation(p, t, n, l)를 바탕으로 각각 정의하고자 한다. 그러나 데이터마트로부터 이들 고객 데이터 집합을 추출하기 위해서는 추가적 변수들이 필요한데 이들은 각각 예측 기간(Prediction Steps), ps 와 추출 대상 기간(Duration), d 그리고 모델 세트 추출 시점(Model Set Start Time), $msst$ 등이 그것이다.

이렇게 예측 기간과 추출 대상 기간이 정해지면 모델 세트는 $msst$ 로부터 $msst+d$ 까지의 기간 동안 서로 다른 상품 클래스 p 개에서 상품을 구매했던 고객 데이터 집합으로 정의할 수 있으며, 스코어 세트는 $t-ps-d$ 부터 $t-ps$ 까지의 기간 동안 서로 다른 상품 클래스 p 개에서 상품을 구매한 고객 데이터 집합으로 정의할 수 있다.

<표 2>는 상품 연관성 분석 수준을 <그림 1>과 같은 상품 분류 체계에서 l 을 1수준으로 선택하고, p 를 1로 그리고 $msst$ 와 d 를 각각 5월과 3개월로 설정하였을 경우, 관련 고객의 데이터마트로부터 모델 세트가 결정되는 과정을 보여주고 있다. 결과적으로 <표 2>상에서 나타난 고객들

중 101, 103, 104 고객만이 이 조건을 만족하므로 {101, 103, 104}가 모델 세트가 되는 것이다.

<표 2> 모델 세트(Model Set) 결정

m	5월	6월	7월	8월	9월	10월	Model Set
101	·	E	·	B		·	T
102	·	·	·	·		·	F
103	A	A	·	·		E	T
104	·	B, D	F	·		C	T

이렇게 선정된 모델 세트에 포함된 고객들에 대하여 고객 속성 정보를 데이터마트로부터 추출하고 이렇게 추출된 고객 속성 정보가 추천 대상 고객 선정 모델은 독립 변수들이 되며 이 때 각 고객의 $msst+d+ps$ 시점의 구매 여부가 종속 변수 즉 고객 선정 모델이 추정하여야 하는 변수가 되는 것이다. 이제 이렇게 추정하여야 하는 종속 변수를 Y 라 하자. 이 때 만약 ps 가 2개월이고 구매한 경우를 1로 구매하지 않은 경우는 0으로 표현한다면 고객 선정 모델을 위한 학습 데이터 집합은 다음의 <표 3>과 같은 데이터 집합이 될 것이다.

<표 3> 학습 데이터(Training Data) 예

m	나이	성별	직업	총 구매액	구매 회수	최근 방문일	...	Y
101	22	M	학생	64	4	0827	...	0
103	36	F	공무원	57	6	1018	...	1
104	23	F	학생	128	10	1104	...	1

이제 이상과 같은 학습 데이터 집합을 좀 더 정형적으로 표현한다면 임의의 N 개의 고객 속성 정보로 이루어진 학습 데이터의 m 번째 고객

에 해당하는 학습 데이터 레코드는 다음처럼 표현할 수 있다.

$$record_m = [X_1, X_2, X_3, X_4, \dots, X_N, Y]$$

$$Y = \begin{cases} 1: msst+d+ps \text{ 시점에 임의의 상품 클래스에 속한 상품을 1개 이상 구매한 고객} \\ 0: 그렇지 않은 고객 \end{cases}$$

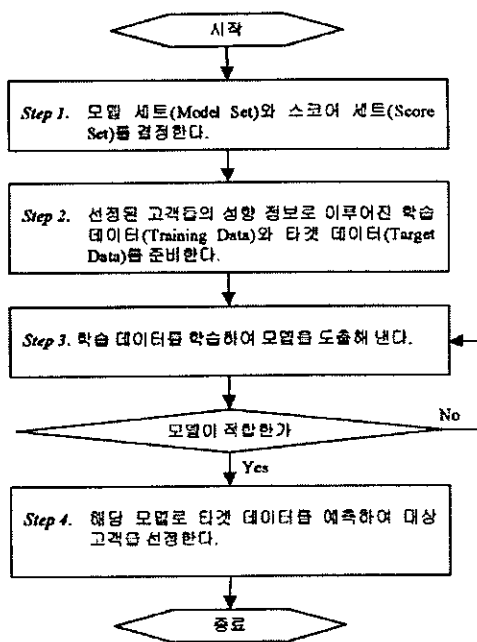
스코어 세트의 결정은 모델 세트와 유사하게 결정되는데 $t-ps-d$ 시점부터 $t-ps$ 시점까지의 기간동안 역시 서로 다른 상품 클래스 p 개에서 상품을 구매한 고객 데이터 집합으로 정의된다. 만약 t 가 12월이라면 $t-ps-d$ 는 $12-2-3=7$ 월이 되며 $t-ps$ 는 12-2로서 10월이 된다. 따라서 이는 7월부터 10월까지 적어도 p 개 즉 현재 예에서는 1개 이상을 구매한 고객들이 스코어 세트를 구성하게 되며 <표 4>의 예에서는 001과 003 고객이 이에 해당되게 되는 것이다. 결과적으로 앞에서 구축된 고객 선정 모델을 이용하여 추천 대상 고객 선정에 적용 될 타겟 데이터 집합은 이들 스코어 세트에 속한 고객들의 속성 정보들을 바탕으로 구성되며 이들 고객의 t 시점에서의 구매 여부가 실제 추정 대상이 되는 것이다.

<표 4> 스코어 세트(Score Set) 결정 예

m	7월	8월	9월	10월	11월	12월	Score Set
001	A	A	E			?	T
002	·		·	·		?	F
003	·	B	·	B		?	T
...

이상에서 정의한 모델 세트와 스코어 세트 그리고 이들 모델 세트와 스코어 세트를 이용하여 준비되는 학습 데이터 집합과 타겟 데이터 집합

을 바탕으로 다음의 <그림 3>에서와 같은 추천 대상 고객 선정 절차를 제안하고자 한다. 이 절차에 따르면 먼저 *ps*, *d*, *msst* 등을 설정함으로써 모델 세트와 스코어 세트를 구성하고 이어 이들 각 집합에 소속된 고객들에 대해 데이터마트를 이용하여 학습 및 타겟 데이터 집합을 준비하게 된다. 다음 단계에서는 학습 데이터 집합을 이용하여 추천 대상 고객 선정을 위한 의사결정 나무 모델을 구축하며 이의 구축의 그 학습 성과를 바탕으로 만족한 수준에 이를 때까지 반복될 수 있다. 만일 만족스러운 추천 대상 고객 선정 모델이 얻어졌다면 이제 스코어 세트에 소속된 고객에 대한 추천 대상 고객 선정 작업이 수행되며 이 과정에서 추천 가능 고객으로 추정된 고객 집합만이 최종 추천 대상으로 선정되어 이들 고객에 적합한 추천 상품을 준비하게 되는 것이다.



<그림 3> 대상 고객 선정 과정

예를 들어 위 <표 4>에서 {001, 003}이 스코어 세트이지만 추천 대상 선정 모델에 의해 003만이 추천 가능 고객으로 추정되었다면 003 고객에 대한 상품 추천만이 발생하게 되는 것이다. 이제 다음 단계에서는 선정된 고객들에 대하여 어떻게 상품 추천이 이루어지는지에 대해 보다 자세히 논의하고자 한다.

3.5 상품 클래스간 연관성 분석 단계

상품 클래스간 연관성 분석 단계란 임의의 기간동안 발생한 고객의 구매 데이터를 바탕으로 앞에서 지정한 1수준에서의 상품 클래스간 연관성을 분석하고 이를 이용하여 각 고객별 연관성 기반 상품 클래스 추천 모델을 수립하는 단계이다. 앞의 예들에서와 같이 $l=1$ 이라 하면 연관성 분석 대상 클래스들은 <그림 1>에서 볼 수 있듯이 A, B, C, D, E, F의 상품 클래스들이 될 것이다. 이러한 연관성 분석을 위한 알고리즘으로는 잘 알려진 Apriori 알고리즘(Agrawal and Srikant, 1994)을 사용하고 있다. 특히 본 연구에서는 구매 데이터외에도 웹로그 분석을 통해 얻어진 장바구니 데이터를 바탕으로 얻어진 연관성 규칙들을 사용하여 인터넷 쇼핑몰 고객의 구매 성향을 보다 더 정확히 반영코자 하였다. 다음에서 이러한 상품 클래스간 연관성 분석을 위한 일련의 단계들을 간략히 소개하고자 한다.

상품 클래스간 연관성 분석 절차

Step 1. 분석 기간을 *msst*시점부터 *t-1*시점까지의 구간으로 설정한다.

Step 2. 분석기간 동안 판매된 상품들을 트랜잭션 분류 1수준에 따라 재분류하고, 구매 데이터를 해당 상품 클래스 수준의 자료로 변

환시킨 뒤 Apriori 알고리즘을 적용하여 연관성 규칙들을 추출한다. 이 과정에서 발견된 규칙 집합을 $Ruleset_p$ 로 표기한다.

Step 3. 분석기간 동안 장바구니에 담긴 상품을 역시 트랜잭션 분류 1수준에 따라 분류하고, 상품 클래스 수준 자료로 변환한 후, Apriori 알고리즘을 적용하여 연관 규칙들을 탐색한다. 일반적으로 구매 데이터에 비해 장바구니 데이터의 신빙성이 약하다고 할 수 있으므로 이 때 적용되는 최소지지도 (Minimum Support) 수준은 Step 2에서 적용한 기준보다 최소한 같거나 높게 설정한다. 이 과정에서 구한 규칙 집합을 $Ruleset_b$ 로 표기한다.

Step 4. $Ruleset_p$ 와 $Ruleset_b$ 의 합집합으로서 최종 $Ruleset_{all}$ 을 구성한다. 이 때, $Ruleset_p$ 와 $Ruleset_b$ 에 동일한 규칙이 존재할 경우, $Ruleset_{all}$ 에는 그 규칙의 신뢰도 값이 큰 경우를 적용한다.

Step 5. $Ruleset_{all}$ 을 바탕으로 고객별 연관성 기반 상품 클래스 추천 모델을 생성한다.

위의 제 5 단계에서의 고객별 연관성 기반 상품 클래스 추천 모델이란 각 고객별로 발견된 연관성 규칙에 의거하여 각 상품 클래스들에 대한 추천 가능성을 표현하고자 하는 것이다. 예를 들어 임의의 고객이 A 상품 클래스를 구매했고 연관성 분석에 의하여 $A \Rightarrow B$ 라는 연관성 규칙이 발견되었다면 이 고객에 대한 추천 가능 상품 클래스는 B라 말할 수 있을 것이다. 그렇다면 만약 이 규칙외에도 $A \Rightarrow C$ 라는 규칙도 발견되었다면 C 상품 클래스 역시 추천 가능 상품 클래스이라고 할 수 있으나 B와 C중 어느 것을 우선적으로 추천해야 할 지에 대한 이러한 기준으로서 가능

한 대안은 바로 각 규칙의 신뢰성이라 할 수 있으며 $A \Rightarrow B$ 의 신뢰성이 $A \Rightarrow C$ 의 신뢰성보다 높다면 B가 C에 대하여 우선적으로 추천되어야 할 것이다.

이상의 맥락에서 본 연구에서는 추천 대상으로 선정된 각 고객별로 추천 가능 상품 클래스와 이들 클래스간의 중요도를 표현하는 연관성 기반 상품 클래스 추천 모델을 벡터 형태로써 아래와 같이 제안하였다. 먼저 임의의 고객 m에 대한 연관성 기반 상품 클래스 추천 모델을 $A^{(m)}$ 라 표기하고자 하며 이 고객이 구매한 상품들의 상품 클래스 집합을 $PurSet_m$ 이라 하고 구매한 상품 클래스들에 의해 연관 규칙을 이용하여 유추할 수 있는 상품 클래스 집합을 $AssoSet_m$ 이라 정의하자. 예를 들어 $PurSet_m$ 이 {A, B}이고 $Ruleset_{all}$ 이 $\{A \Rightarrow C, B \Rightarrow D\}$ 라면 $AssoSet_m$ 은 {C, D}가 됨을 의미한다. 이제 이들 두 정의를 바탕으로 $A^{(m)}$ 은 다음과 같이 정형적으로 정의할 수 있다.

$$A^{(m)} = [A_{m1}, A_{m2}, \dots, A_{ms}, \dots, A_{mS}]^T, m = 1, \dots, M$$

$$A_{ms} = \begin{cases} conf(s) & : \text{if } s \in AssoSet_m - PurSet_m \\ 0 & : \text{그 외의 경우} \end{cases}$$

여기서 $conf(s)$ 는 $AssoSet_m$ 에 속하는 임의의 상품 클래스 s가 도출된 연관 규칙의 신뢰도 (Confidence)를 의미하며, 만약 고객 m에 대하여 상품 클래스 s를 결과부로 갖는 연관 규칙이 여러 개일 경우, 연관 상품 클래스 s의 신뢰도는 연관 규칙들 중에 가능한 가장 큰 신뢰도 값으로 설정하게 된다. 결과적으로 $A^{(m)}$ 벡터의 각 값들은 연관 규칙으로부터 유추할 수 있는 고객 m의 해당 상품 클래스에 대한 구매 가능성을 말한다 하겠다.

예를 들어 이상의 상품 클래스간 연관성 분석 절차를 소개한다면 먼저 <표 5>에서와 같이 고객의 구매 및 장바구니 자료로부터 상품 클래스간 연관 규칙들이 생성되었다고 가정하자. 이 표를 통해 $Ruleset_p$ 와 $Ruleset_b$ 는 각각 최소지지도를 0.05와 0.1로서 생성되었음을 알 수 있다.

<표 5> 상품 클래스간 연관성 규칙 생성 예

Rule Set	상품 클래스간 연관 규칙
$Ruleset_p$ ($ms=0.05$)	A, E \Rightarrow C(0.8), B \Rightarrow E(0.4), B \Rightarrow F(0.5), C \Rightarrow A(0.2)
$Ruleset_b$ ($ms=0.1$)	A, E \Rightarrow C(0.6), B \Rightarrow E(0.8), C \Rightarrow A(0.7), C \Rightarrow F(0.4), A \Rightarrow D(0.6)
$Ruleset_{all}$	A, E \Rightarrow C(0.8), A \Rightarrow D(0.6), B \Rightarrow E(0.8), B \Rightarrow F(0.5), C \Rightarrow A(0.7), C \Rightarrow F(0.4)

이제 아래의 <표 6>에서와 같이 001, 003, 012, 017 고객들이 추천 대상 고객으로 선정되었고 이들의 과거 구매 이력이 각각 표에서의 $PurSet_m$ 과 같다면 위의 <표 5>에서의 $Ruleset_{all}$ 의 연관 규칙들을 이용하여 $AssoSet_m$ 들과 같은 추천 가능 상품 클래스와 이에 대한 추천 가능성들을 도출할 수 있는 것이다.

<표 6> 고객별 $AssoSet$ 의 도출

m	$PurSet_m$	$AssoSet_m$
001	A, E	C(0.8), D(0.6)
003	B	E(0.8), F(0.3)
012	B, C, E	A(0.7), F(0.4)
017	B, D	E(0.8), F(0.3)

이제 우리는 앞에서의 연관성 기반 상품 클래스 추천 모델 $A^{(m)}$ 에 대한 정의와 <표 6>에서와 같은 $PurSet_m$ 과 $AssoSet_m$ 를 이용하여 각 고객

별 연관성 기반 상품 클래스 추천 모델을 도출할 수 있다. 예를 들어 001 고객에 대한 연관성 기반 상품 클래스 추천 모델 $A^{(001)}$ 은 <표 6>에서의 첫번째 행을 이용하여 [0, 0, 0.8, 0.6, 0, 0]로 설정할 수 있는 것이다.

3.6 고객 상품 선호도 분석 단계

앞 절에서 다룬 연관성 기반 상품 클래스 추천 모델은 고객의 구매 및 장바구니 데이터를 분석함으로써 도출된 연관 규칙을 기반으로 구성된 상품 클래스 추천 모델이라 할 수 있다. 따라서 이는 각 고객의 성향에 근거한 것이라기 보다는 전체 고객의 평균적 성향에 의거한 상품의 추천 정보라 할 수 있는 것이다. 앞서서도 밝힌 바와 같이 본 연구에서는 상품 추천의 효과를 극대화시키기 위하여 이러한 전체 고객의 평균적 성향 뿐만 아니라 고객 개인의 성향도 상품 추천에 반영하고자 하며 이를 위해 각 고객별 상품 클래스 선호도 모델(Product Class Preference Model)을 제안하고자 한다.

이와 같은 고객의 상품 클래스 선호도 모델 역시 벡터로서 표현하고자 하며 벡터의 각 값은 고객이 해당 상품 클래스에 대해 어느 정도의 구매 선호도를 가지고 있는지를 측정하고자 한다. 본 연구에서는 우선 구매 선호도의 측정을 위해 고객이 인터넷 쇼핑몰에서 상품을 구매하는 동안 발생하는 두가지 행동 패턴 정보를 분석하고 이를 바탕으로 측정 방법을 제안하고자 한다. 먼저 이들 두가지 행동 패턴은 각각 고객의 특정 상품 클래스의 상품들에 대한 클릭 횟수와 장바구니 등장 횟수로서 이들은 각기 해당 고객의 특정 상품 클래스에 대한 구매하지는 않았으나 그 구매 선호도를 반영한다고 볼 수 있다.

임의의 고객 m 의 상품 클래스 선호도 모델을 $C^{(m)}$ 이라 하고 이 고객의 상품 클래스 s 에 속하는 상품들에 대한 배너 총 클릭 수를 V_{ms} , 장바구니에서의 출현 빈도를 B_{ms} 라 한다면 $C^{(m)}$ 은 다음과 같이 정의할 수 있다.

$$C^{(m)} = [C_{m1}, \dots, C_{ms}, \dots, C_{mS}]^T, m = 1, \dots, M, \text{ where}$$

$$C_{ms} = \frac{\hat{V}_{ms} + \hat{B}_{ms}}{2}$$

여기서 \hat{V}_{ms} 는 전체 상품 클래스의 수를 S 개라 할 때, 아래의 식과 같이 V_{ms} 를 전체 상품 클래스에 대한 클릭 수에 대해 정규화 시킨 값을 의미하고, \hat{B}_{ms} 역시 B_{ms} 를 전체 장바구니 출현 빈도에 대하여 s 상품 클래스에 대한 빈도를 정규화 시킨 값을 의미한다.

$$\hat{V}_{ms} = \frac{V_{ms}}{\sum_{s'=1}^S V_{ms'}}, \hat{B}_{ms} = \frac{B_{ms}}{\sum_{s'=1}^S B_{ms'}}$$

다음의 <표 7>은 앞의 예들에서 선택된 추천 대상 고객 001, 003, 012, 107 고객들에 대한 상품 클래스 선호도 모델의 예를 보여주고 있다.

<표 7> 고객의 상품 클래스 선호도 모델 예

m	A	B	C	D	E	F
001	0.2	0	0.15	0	0.2	0
003	0	0.2	0.1	0	0.1	0.08
012	0.2	0	0.1	0	0.15	0.22
017	0	0.12	0	0.2	0.1	0.15

3.7 상품 추천 단계

이상에서 우리는 각 고객별 연관성 기반 상품

클래스 추천 모델 $A^{(m)}$ 와 상품 클래스 선호도 모델 $C^{(m)}$ 을 도출하였다. 이제 $A^{(m)}$ 에 반영된 평균적 구매 가능성과 $C^{(m)}$ 에 반영된 고객별 개인 상품 클래스 선호도를 바탕으로 각 고객별 상품 클래스별 추천 가능성을 평가하고자 한다. 이들 두 모델들을 바탕으로 각 고객에 대한 특정 상품 클래스의 추천 가능성을 통합 평가하기 위하여 우리는 매칭 스코어(Matching Score)라는 다음과 같은 추천 가능성 측정 방법을 제안하였다. 고객 m 의 상품 클래스 s 에 대한 매칭 스코어(Matching Score)를 $MS(m, s)$ 라 표기하고 이에 대한 정의는 다음과 같다.

$$MS(m, s) = \frac{2 \times A_s^{(m)} \times C_s^{(m)}}{A_s^{(m)} + C_s^{(m)}}$$

이 공식의 의미는 직관적으로 고객의 상품 선호도도 높고 평균적 구매 가능성도 높다면 해당 상품 클래스 s 의 고객 m 에 대한 추천가능성도 높다고 해석할 수 있는 것이다. 이와 같은 매칭 스코어 평가 방법에 따라 <표 6>과 <표 7>로부터 각 고객별 및 각 상품 클래스별로 아래의 <표 8>과 같은 매칭 스코어 테이블을 얻을 수 있다.

<표 8> 매칭 스코어 테이블 예

m	A	B	C	D	E	F
001	0	0	.253	0	0	0
003	0	0	0	0	.178	.126
012	.311	0	0	0	.0	.284
017	0	0	0	0	.178	.2

이제 이와 같은 매칭 스코어 테이블을 이용하여 각 고객에 대한 실질적 상품 추천 절차를 다음과 같이 제안하고자 한다.

상품 추천 절차

- Step 1.** 추천 담당자가 추천 상품 클래스의 수, nc 를 결정한다.
- Step 2.** 각 고객 별로 상위 매칭 스코어(Matching Score) 값을 갖는 $nc (< n)$ 개의 상품 클래스를 결정한다.
- Step 3.** 각 고객별로 선정된 각 상품 클래스별로 상위 Click-to-Buy Rate를 갖는 n/nc 개의 상품 클래스의 단품들을 선정하여 추천한다.

특히 위의 제 3 단계에서의 추천 상품으로서의 단품에 대한 결정은 각 단품별 Click-to-Buy Rate에 의거하여 상위 단품들을 추천함으로써 실질적인 추천 효과를 추가적으로 제고하고자 한 것이다. 예를 들어 아래의 <표 9>와 같이 각 상품 클래스별로 소속 단품들이 존재하고 이들 각 단품들의 Click-to-Buy Rate들이 이와 같다고 가정하자.

<표 9> 단품별 Click-to-Buy Rate 예

상품 클래스	A	B	C	D	E	F
단품 상품	A ₁ (.01)	B ₁ (.02)	C ₁ (.03)	D ₁ (.01)	E ₁ (.02)	F ₁ (.00)
	A ₂ (.05)	B ₂ (.02)	C ₂ (.02)	D ₂ (.01)	E ₂ (.01)	F ₂ (.015)
	A ₃ (.02)	B ₃ (.04)	C ₃ (.00)	D ₃ (.00)	E ₃ (.01)	F ₃ (.01)
	A ₄ (.03)		C ₄ (.03)	D ₄ (.02)		F ₄ (.01)
		C ₅ (.04)				

이 예에서 만약 인터넷 쇼핑몰의 상품 추천 담당자가 nc 를 2로 설정하였고 앞에 예제 추천 문제 상황을 그대로 이용한다면 n 이 2이므로 각 상품 클래스당 추천해야할 단품의 수는 n/nc 로서 1이 되며 따라서 고객 003의 경우 <표 8>에 의거하여 E와 F 상품 클래스가 추천 상품 클래스로 선정되며 각 클래스별로는 가장 높은 Click-to-

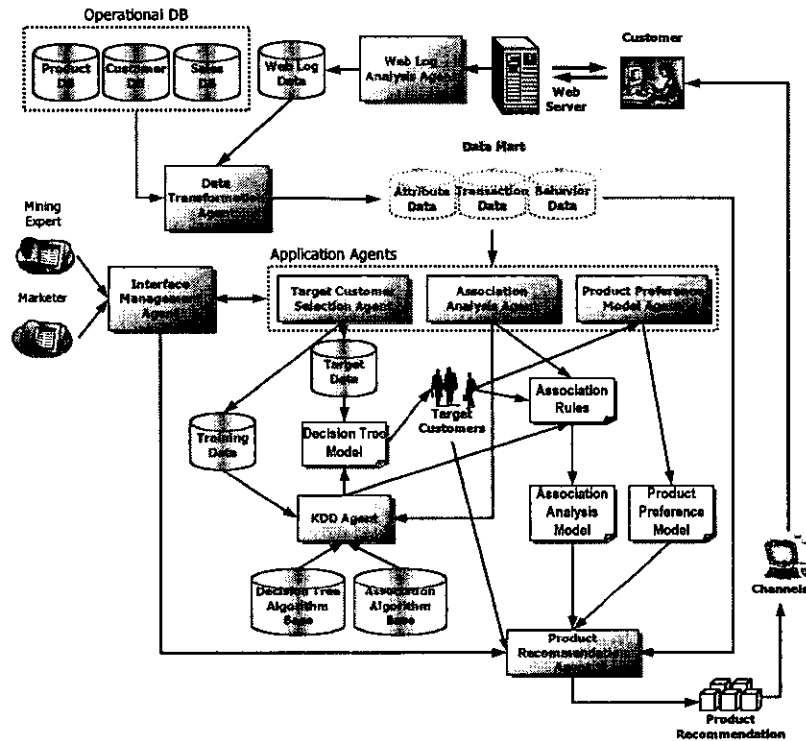
Buy Rate를 갖는 상품들이 <표 9>를 이용하여 각각 E1과 F2로 선정되는 것이다. 결과적으로 003 고객에게는 E1과 F2를 추천하게 된다. 다음의 <표 10>은 앞에서 예로든 모든 추천 대상 고객에 대하여 선정된 이러한 추천 상품들의 예를 보여주고 있다. 특기한 것은 001 고객의 경우 2개의 상품을 추천해야 함에도 불구하고 추천 대상 클래스가 <표 8>에서와 같이 1개밖에 유의하지 않아 동일 클래스 내에서 추천 대상 상품을 추출하게 되었다.

<표 10> 고객별 추천 상품 예

m	구매 상품 클래스	추천 상품 클래스	추천 상품
001	A, E	C, D	C5, C4
003	B	E, F	E1, F2
012	B, C, E	A, F	A2, F2
017	B, D	E, F	E1, F2

4. 에이전트 및 지식탐사기술 기반 지능형 인터넷 쇼핑몰을 위한 상품 추천 시스템

본 장에서는 이상에서 다룬 인터넷 쇼핑몰에서의 상품 추천 방법론을 기반으로 에이전트 및 지식탐사기술을 이용하여 지능형 인터넷 쇼핑몰을 위한 상품 추천 지원 시스템을 구조적 설계와 이의 각 구성 에이전트들에 대하여 소개하고자 한다. 먼저 다음의 <그림 4>는 본 연구에서 제안하고자 하는 지능형 인터넷 쇼핑몰을 위한 에이전트 기반 상품 추천 지원 시스템의 개략적 구조를 보여주고 있다.



<그림 4> 지능형 인터넷 쇼핑물을 위한 에이전트 기반 상품 추천 지원 시스템의 구조도

위 그림에서 볼 수 있듯이 본 시스템은 크게 8개의 에이전트들과 관련된 데이터베이스 또는 데이터마트로 구성되어 있다. 이제 이들 8개의 에이전트들의 기능과 역할에 대해 소개함으로써 전체적인 시스템이 앞에서 다룬 상품 추천 지원 방법론을 어떻게 구현하고 있는지에 대해 설명하고자 한다.

웹로그 분석 에이전트(Web Log Analysis Agent): 정기적으로 웹 서버에 존재하는 접근 로그 파일(Access Log File), 참조자 로그 파일(Referer Log File), 에이전트 로그 파일(Agent Log File) 등을 파싱 및 분석을 통하여 다른 운영 데이터베

이스와 마찬가지로의 접근 및 분석이 가능하도록 웹로그 데이터베이스를 유지 관리한다.

데이터 변환 에이전트(Data Transformation Agent): 상품 추천 문제의 해결에 필요한 데이터마트를 유지 관리하기 위해 정기적으로 운영 데이터베이스(Operational Database)와 웹로그 데이터베이스(Web Log Database)로부터 자료를 추출(Extraction) 및 정제(Cleansing) 과정등을 일차적으로 수행하고 이차적으로는 이들 데이터들을 상품 추천 방법론의 적용에 적합한 형태로 변환시키는 과정을 수행함으로써 데이터마트의 적시성과 데이터 사용 용이성을 유지 관리하게 된다.

지식탐사 에이전트(KDD Agent): 지식탐사 에이전트는 상품 추천 문제와 관련된 지식 탐사 알고리즘들을 관리하며 응용 문제 에이전트들의 지식탐사 요청에 대하여 지식 탐사 작업을 수행하고 그 결과로 생성되는 지식이나 모델들을 제공하게 된다. 역시 생성된 지식이나 모델들에 대한 관리를 수행하며 생성된 모델이나 지식을 이용한 추론이나 예측 서비스도 요청에 의해 지원하는 역할을 담당하게 된다. 이러한 지식 탐사 에이전트를 독립적으로 운영함으로써 탐사 방법론의 대체나 탐사 기법의 조정으로부터의 전체 시스템의 안정성을 확보할 수 있는 것이다.

목표 고객 선정 에이전트(Target Customer Selection Agent): 응용 에이전트 군 중 첫번째에 해당하는 에이전트로 3.4절에서 소개한 추천 대상 고객 선정 단계를 수행하는 역할을 담당하고 있다. 인터페이스 관리 에이전트를 통하여 지식 탐사 전문가나 마케터와의 의사 전달이 가능하며 이를 통해 설정된 업무를 수행하기 위해 데이터마트를 통해서 필요한 고객 데이터를 획득하며 지식탐사 에이전트와의 협조를 통해 필요한 고객 선정 모델을 수립하게 된다. 본 연구에서는 의사결정나무 기법을 이용하므로 결과적으로 의사결정나무 모델을 확보하고 이를 이용하여 설정 업무에서 제시한 기간에 대한 상품 추천 목표 고객을 선정하게 되는 것이다.

연관성 분석 에이전트(Association Analysis Agent): 연관성 분석 에이전트는 응용 에이전트 군 중 두번째 에이전트로서 3.5절의 상품 클래스 간 연관성 분석 단계를 수행하는 역할을 맡고 있다. 이를 위해 역시 데이터마트로부터 고객 구매 데이터를 획득하고 지식탐사 에이전트의 협조를 통하여 연관성 규칙 탐사 과정을 수행함으로써

상품 연관성 규칙들을 생성하게 된다. 이어 목표 고객 선정 에이전트에 의해 생성된 목표 고객들에 대한 연관성 기반 상품 클래스 추천 모델을 생성된 연관성 규칙을 이용하여 제공하고 있다.

상품 선호도 모델 에이전트(Product Preference Model Agent): 응용 에이전트 군의 마지막 에이전트로서 3.6절의 고객 상품 선호도 분석을 담당하고 있다. 역시 목표 고객 선정 에이전트에 의해서 결정된 목표 고객들에 대해서 과거 구매 데이터를 이용 각 고객별 상품 선호도 모델을 생성하게 된다.

상품 추천 에이전트(Product Recommendation Agent): 상품 추천 에이전트는 데이터마트로부터의 상품 관련 데이터와 연관성 분석 에이전트와 상품 선호도 모델 에이전트에 의해서 작성된 연관성 기반 상품 클래스 추천 모델과 상품 선호도 모델을 이용하여 각 고객별 최종 추천 상품 리스트를 작성하게 된다. 이 때 역시 마케터로부터의 추천 상품 클래스 수에 대한 정보가 인터페이스 관리 에이전트에 의해 전달되게 된다.

인터페이스 관리 에이전트(Interface Management Agent): 인터페이스 관리 에이전트는 본 시스템의 사용자 즉 마케팅 담당자나 지식탐사 전문가 등과 에이전트들 사이의 인터페이스를 생성하고 상호 정보의 교환이 가능하도록 해주며 이로 인해 사용자는 필요한 통제나 기초 정보를 각 에이전트들에 전달할 수 있으며, 에이전트들은 업무의 결과나 필요한 보고를 담당자에게 전달할 수 있게 되는 것이다.

현재 이상의 각 에이전트들과 필요한 데이터 마트 등에 대한 프로토타입 시스템은 자바를 기반으로 개발되어 있으며, 상품 추천 방법론과 개

발된 시스템의 성과는 다음에서 사례를 바탕으로 경험적 분석을 해보고자 한다.

5. 사례연구

이상에서 소개한 지능형 인터넷 쇼핑물을 위한 에이전트 기반 상품 추천 지원 시스템은 국내 유명 인터넷 쇼핑물 중 하나인 C사이트에 적용시켜 보았다. 이 사이트는 주로 화장품 판매를 목적으로 하고 있으며, 데이터베이스 시스템으로 MS SQL Server 7.0을 사용하고 있으며, 웹 서버로는 IIS 5.0을 사용하고 있다. 이 사이트의 등록 고객은 36만 여명에 달하며, 판매하는 상품은 총 3천 여개에 이른다. 또한 이들 상품들은 대분류-중분류-소분류로 3개의 제품 분류 수준을 가지고 있으며, 최상위 수준에서는 10개의 상품 클래스로 관리하며 다시 이들 10개 클래스가 다음 수준에서는 20개의 상품 클래스로 마지막 3 수준에서 100여개의 상품 클래스로 관리되고 있다.

먼저 본 시스템의 적용을 위해 해당 사이트의 2001년 1월부터 5월까지의 웹로그 데이터와 운영 데이터베이스를 기반으로 운영 실험을 실시하였으며, 상품 추천 과정의 분석 단위로 3 수준의 상품 클래스, 즉 100여개의 상품 클래스들을 이용하여 상품 추천 방법론을 적용하였다. 구체적 상품 추천을 수행하기 위해 마케팅 담당자는 적용 대상 문제를 인터페이스 관리 에이전트를 통하여 Recommendation(1, 2001/6, 3, 3)로 초기 설정하였으며, 역시 동일한 에이전트를 통하여 지식탐사 담당자에 의해 모델 세트 시작 시점, $msst$ 는 2001년 1월로 모델 세트 기간 d 는 2개월로 그리고 예측 스텝 기간, ps 는 1개월로 추가 설정되었다.

결과적으로 이러한 설정으로 인해 목표 고객

선정 에이전트는 데이터마트로부터 2001년 1월에서 3월 까지의 기간 동안 최소 1개 이상의 상품을 구매한 고객들을 선정하였고 이들 선정된 고객들에 대하여 4월에 상품 구매를 했던 고객은 추천 대상 평가 값을 1로 그렇지 않은 경우에는 0으로 설정하였다. 추천 대상 평가를 위한 독립 변수들로는 마케팅 담당자와 지식탐사 담당자의 협의를 통하여 나이, 직업, 최근 방문일, 총 구매액, 구매 회수 등으로 결정하였으며 이와 같은 과정을 통하여 의사결정나무 모델을 구성할 학습 데이터를 준비하였다. 준비된 학습 데이터를 바탕으로 목표 고객 선정을 위한 의사결정나무 모델이 도출되었으며, 이젠 2001년 3월부터 5월까지 최소 1개 이상의 상품 구매를 한 고객들을 대상으로 도출된 의사결정나무 모델을 이용하여 추천 대상 고객들을 선정하였다.

다음으로 상품 연관성 분석 에이전트는 지정된 기간 2001년 1월에서 5월까지의 구매 데이터와 장바구니 데이터를 이용하여 상품 클래스간 연관 규칙을 탐사하였으며 이 때, 구매 상품간 연관 규칙에 대한 최소 지지도는 0.02로 장바구니 상품간의 연관 규칙에 대한 최소 지지도는 0.05로 각각 설정하였다. <표 11>은 이와 같은 연관 규칙 탐사를 통해 이 사례에서 도출된 실제 연관 규칙들의 일부를 예시하고 있다.

<표 11> 발견된 연관 규칙 예

No.	조건부	결과부	지지도	신뢰도
1	팩	폼클렌징	0.053	0.5
2	기초세트	화이트닝	0.153	0.65
3	선크림	바다로션	0.07	0.54
4	투웨이케익	에센스	0.025	0.46
...
52	트리트먼트	스타일링보조제	0.03	0.8

이렇게 도출된 연관 규칙들은 위 표에서 볼 수 있는 바와 같이 총 52개이었으며 이를 이용하여 선정된 추천 대상 고객별로 상품 연관성 기반 상품 추천 모델을 도출하였다. 이외에도 상품 선호도 모델 에이전트는 역시 2001년 1월부터 5월까지의 구매 이력을 바탕으로 각 고객별 상품 선호도 모델을 구성하였다.

마케팅 담당자는 역시 인터페이스 관리 에이전트를 통해 상품 추천 에이전트에 대하여 추천 상품의 클래스를 3개로 하도록 설정하였으며 이에 따라 상품 추천 에이전트는 각 고객별로 매칭스코어 관점에서 상위 상품 클래스 3개를 선정하고 역시 각 선정된 클래스 별로 Click-to-Buy Rate가 가장 높은 상품을 하나씩 추천하였다. 아래의 <그림 5>는 이러한 과정을 통해 상품 추천 에이전트에 의해 실제 각 고객별로 추천된 상품 상황을 보여주고 있으며 실질적으로 이 화면은 상품 추천 에이전트가 도출한 자료를 인터페이스 관리 에이전트에 의해 웹 상 화면으로 변환되어 사용자에게 도시되고 있는 것이다.

<그림 5> 상품 추천 결과 보기 화면 예

6. 결론

빠르게 성장하고 있는 전자상거래 환경에서 인터넷 쇼핑몰 기업들은 단순히 고객을 상품 판매의 대상으로서가 아닌 비즈니스라는 매개체를 통하여 관계를 지속시켜야만 하는 대상으로 인식하게 되었고 이에 따라 고객의 다양하면서도 변화하는 요구에 어느 정도 부응할 수 있는지가 기업 경쟁력의 척도이며 유일한 생존 수단이라 보여 진다. 본 연구에서는 이러한 고객 맞춤화 시대에 있어서 실질적으로 인터넷 쇼핑몰 기업의 의사 결정 지원과 고객과의 관계 강화의 수단으로서 상품 추천 지원 방법론을 제안하였으며 이를 구현하기 위하여 에이전트와 지식탐사기법에 기반을 둔 지능형 인터넷 쇼핑몰 시스템을 설계하였다.

이러한 관점 하에서 본 연구에서 제안한 방법론과 시스템 설계의 실질적 의의를 다음과 같이 정리할 수 있겠다.

첫째, 대부분의 기존 상품 추천 방법론들은 고객의 구매 이력 정보에 대한 의존도가 높음으로 인해 인터넷 쇼핑몰과 같은 웹 사이트에서 웹 로그를 통해 추가적 획득가능한 고객의 행태 정보를 활용하기가 어려웠으며 본 방법론에서는 이와 같은 고객의 구매 이력 정보는 물론 웹 로그 정보까지 상품 추천에 반영함으로써 보다 정확한 고객의 상품 선호도 파악이 가능해졌으며 이에 따라 보다 효과적인 상품 추천을 가능하게 하였다.

둘째, 역시 대부분의 상용화된 기존 방법론들은 항상 모든 고객을 대상으로 상품 추천을 시도하고 있으며 이는 곧 많은 경우에 스팸 메일과 같은 부적합한 추천에 의한 부정적 효과를 가져오게 된다. 본 연구에서는 이를 최소화하기 위하

여 상품 추천에 있어 먼저 추천 가능한 고객을 추출하는 예측 모델링 기법을 이용함으로써 이러한 부정적 효과를 최소화하고 있다.

셋째, 상품간 연관성과 동시에 고객의 개인적 취향을 고려할 수 있도록하기 위해 연관성 기반 추천 상품 모델과 상품 선호도 모델을 고안하고 이를 통합 고려할 수 있는 추천 방법론을 제안함으로써 추천 효과를 극대화시키고 있으며 또한 추천에 대한 응답율을 높이기 위해 상품별 클릭 스트림까지도 반영하는 추천 방식을 채택하였다.

마지막으로 이들 방법론은 효과적으로 구현하기 위해 에이전트 기법을 이용한 지능형 인터넷 쇼핑몰 지원 시스템을 설계하였고 프로토타입 시스템을 구현하였으며 실제 인터넷 쇼핑몰 사이트에서 운영적 및 추천의 기능적 측면의 경험적 타당성을 검증하였다.

본 연구에 대한 경제적 타당성 즉 추천에 의한 매출이나 순이익의 증대 정도 또는 추천에 대한 응답율 등의 관점에서는 현재 성과 분석이 진행 중이다. 현재까지는 매우 고무적인 중간 결과가 나온 것으로 판단하고 있으며 또한 이차적으로는 여타 추천 방법론들 즉, 협업 필터링 상품 추천 방법론 및 연관 규칙 기반 상품 추천 방법론들과 성과 비교가 필요하다고 판단하고 있으며 본 연구의 향후 과제로서 준비 중에 있다.

참고문헌

- Adomavicius, G. and A. Tuzhilin, "Expert-driven Validation of Rule-Based User Models in Personalization Applications", Proceedings of the Fifth ACM SIGKDD International Conference on knowledge Discovery and Data Mining(KDD-99), 1999.
- Agrawal, R., T. Imielinski, and A. Swami: "Mining Association Rules between Sets of Items in Large Database", Proceedings of the ACM SIGMOD Conference on Management of Data, Washington, D.C., May 1993, pp. 207-216.
- Agrawal, R. and R. Srikant: "Fast Algorithms for Mining Association Rules", Proceedings of the 20th VLDB Conference, Santiago, Chile, September 1994.
- Berson, A. and S. J. Smith, Data Warehousing, Data Mining and OLAP, McGraw Hill, 1997.
- Borchers, A., J. Herlocker, J. Konstan, and J. Riedl, "Ganging up on Information Overload", IEEE Computer, 31(4), April 1998, pp. 106-108.
- Buchner, A. and M. D. Mulvenna, "Discovering internet marketing intelligence through online analytical web usage mining", SIGMOD Record, 27(4), 1998, pp. 54-61.
- Cooley, R., B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", Knowledge and Information Systems, 1999
- Han, J. and Y. Fu, "Mining Multiple-Level Association Rules in Large Databases", IEEE Transaction on Knowledge and Data Engineering, Vol.11, No.5, September/October 1999.
- Klemettinen, M., H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo, "Finding Interesting Rules from Large Sets of Discovered Association Rules", Proceedings of Third Int'l Conf. Information and Knowledge Management, Gaithersburg, Md., Nov. 1994, pp. 401-408
- Konstan, J., B. Müller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News", Communications of the ACM, 40(3), 1997, pp. 77-87.

- Lawrence, R. D., G. S. Almasi, V. Kotlyar, M. S. Viveros, and S. S. Duri, "Personalization of supermarket product recommendations", Technical Report RC 21792, IBM, 2000.
- Lee, J. and M. Podlaseck, "Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising", Working Paper, IBM T.J. Watson Research Center, 2000.
- Quinlan, J. R., "Induction of Decision Trees", Machine Learning, 1, 1986, pp. 81-106.
- Quinlan, J. R., C4.5 Programs for Machine Learning, Morgan Kaufmann 1993.
- Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", Proceedings of CSCW '94, Chapel Hill, NC, 1994, pp.175-186.
- Resnick, P. and H. R. Varian, "Recommender Systems", Communications of the ACM, 40(3), 1997, pp. 56-58.
- Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommendation Algorithm for E-Commerce", Proceedings of EC'00 Conference, Minneapolis, Minnesota, October 17-20, 2000.
- Wolf, J., C. Aggarwal, K-L. Wu, and P. Yu, "Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering", Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, San Diego, CA., 1999, pp 201-212.
- Yuan, S. and W. Chang, "Mixed-initiative Synthesized Learning Approach for Web-based CRM", Expert Systems with Applications, 20, 2001, pp. 187-200.

Abstract

Development of Intelligent Internet Shopping Mall Supporting Tool Based on Software Agents and Knowledge Discovery Technology

Jae-kyeong Kim*, Woo-ju Kim**
Yoon-ho Cho***, Je-ran Kim****

Nowadays, product recommendation is one of the important issues regarding both CRM and Internet shopping mall. Generally, a recommendation system tracks past actions of a group of users to make a recommendation to individual members of the group. The computer-mediated marketing and commerce have grown rapidly and thereby automatic recommendation methodologies have got great attentions. But the researches and commercial tools for product recommendation so far, still have many aspects that merit further considerations. To supplement those aspects, we devise a recommendation methodology by which we can get further recommendation effectiveness when applied to Internet shopping mall. The suggested methodology is based on web log information, product taxonomy, association rule mining, and decision tree learning. To implement this, we also design an intelligent Internet shopping mall support system based on agent technology and develop it as a prototype system. We applied this methodology and the prototype system to a leading Korean Internet shopping mall and provide some experimental results. Through the experiment, we found that the suggested methodology can perform recommendation tasks both effectively and efficiently in real world problems. Its systematic validity issues are also discussed.

Key words: Product Recommendation, Internet Shopping Mall, Web Log Analysis, Data Mining, Agent Technology

* School of Business Administration, Kyung Hee University
** Department of Industrial Engineering, ChonBuk National University
*** Department of Internet Information Systems, Dongyang Technical College
**** OpenTech Co., Ltd.