

# 연관규칙기법과 분류모형을 결합한 상품 추천 시스템: G 인터넷 쇼핑몰의 사례

## The Product Recommender System Combining Association Rules and Classification Models: The Case of G Internet Shopping Mall

안 현 철 (Hyunchul Ahn)

한국과학기술원 테크노경영대학원 박사과정

한 인 구 (Ingo Han)

한국과학기술원 테크노경영대학원 교수

김 경 재 (Kyoung-Jae Kim)

동국대학교 경영대학 정보관리학과 교수

### 요 약

오늘날 인터넷이 확산되어감에 따라, e-CRM에 대한 관심이 증대되고 있다. 그 중에서도 특히 ‘추천 시스템’은 e-CRM의 여러 응용분야 중에서도 실무적으로 그리고 학문적으로 가장 활발하게 연구되고 있는 분야 중 하나다. 추천을 위한 여러가지 방법들 중에서, 지금까지 주류를 이뤄온 방법들은 협동 필터링(Collaborative Filtering) 기법과 내용 기반(Content-Based) 접근법이다. 그러나 이러한 기존 방법들은 몇 가지 태생적인 한계점으로 인해 고객의 구매 이력이 많지 않은 중소형 인터넷 쇼핑몰에 적용하기 어렵다는 단점이 있다. 이에, 본 연구에서는 고객의 인구통계 및 구매정보에 2가지 데이터마이닝 기법들(연관 관계 기법과 분류 기법)을 적용하고, 이 결과를 조정 에이전트를 통해 결합하는 형태의 새로운 추천 시스템의 모형과 시스템 구조 체계를 제안한다. 제안된 연구 모형의 유용성을 검증하기 위해, 본 연구에서는 실제 사례에 적용한 웹 기반 프로토타입을 개발, 활용하였다. 프로토타입의 유용성을 실제 사용자들로부터 설문을 통해 조사해 본 결과, 본 연구에서 제안한 추천모형이 생성한 맞춤 정보가 사용자들에게 매우 유익하게 인지됨을 확인하였다.

**키워드 :** 고객관계관리, 추천시스템, 데이터마이닝, 연관관계규칙, 지능형 에이전트

### I. 서 론

최근 인터넷의 확산과 함께 이른바 ‘대중 맞춤(mass customization)’이 기업 경쟁력 강화에 있어 중요한 이슈로 작용함에 따라, 인터넷을 기반으로 한 고객관계관리, 즉 e-CRM에 대한 세간의 관심이 크게 높아지고 있다. 특히 e-CRM

의 여러 분야 중에서도 특정 고객에게 그가 관심을 가질만한 제품이나 서비스를 추천해 주는 ‘추천 시스템(recommender system)’은 실무적으로 가장 널리 이용되는 e-CRM의 수단으로서, 이미 아마존(Amazon), 시디나우(CD Now), 제이씨페니(JC Penny) 등 해외의 유수 사이트는 물론 삼성몰, 한솔CS클럽 등 국내의 일부 쇼핑몰

에서도 널리 적용되고 있다.

현재 추천 시스템의 알고리즘과 관련하여 학계에서 진척되고 있는 연구의 방향은 크게 2가지 형태로 이루어지고 있다. 우선 첫 번째 접근법은 ‘내용 기반(content-based) 필터링 방법’이다. 이것은 방문객에게 그 방문객이 과거에 선호했던 아이템과 유사한 성격을 가진 아이템을 추천하는 방식으로서, 아이템과 아이템 사이의 연관성을 기반으로 추천을 하는 방식이다. 또 다른 접근법은 ‘협동 필터링(collaborative filtering) 접근법’이다. 이것은 앞서 내용 기반 접근법이 아이템간의 연관성을 고려하는 것과 달리, 사용자 간의 연관성을 기반으로 추천하는 방식인데, 특정 사용자와 유사한 성격을 가진 다른 사용자들이 선호한 것으로 밝혀진 아이템을 추천하는 방식이다(Funakoshi and Ohguro, 2000; Wang *et al.*, 2004).

이러한 두 가지 방법론은 물론 각각 여러 장점을 갖고 있는 것이 사실이지만, 다음과 같은 한계점을 가지고 있다. 우선 내용기반 필터링 알고리즘의 경우, 유사한 아이템을 규명하기 위해서는 각 개별 아이템의 특성을 명확하게 추출해 내야 하는데, 이것이 ‘사람’에 의존해야 하기 때문에 제대로 이루어지지 않는 한계점이 있다. 또한 추천의 결과가 주로 사용자가 이미 구매를 했거나, 평가를 내린 아이템과 관련된 것으로만 결정되는 ‘과도한 특수화(overspecialization)’ 문제도 존재한다. 한편 협동 필터링 방법론은 우선 사용자 및 구매 데이터가 어느 정도 축적되어야 방문객과 유사한 성격을 가진 고객군을 찾아내어 추천을 할 수 있다는 한계점이 있다. 또한 협동 필터링 방법론은 고객 레코드의 규모가 커질수록 유사성을 계산하기 위한 연산 처리가 기하급수적으로 늘어나는 ‘확장성 문제(scalability problem)’가 발생하여, 고객 데이터가 방대한 경우 많은 컴퓨터의 연산 용량을 필요로 하는 문제점이 있다. 아울러 아무리 풍부한 고객 데이터를 보유하고 있다고 하더라도, 그 고객이 구

매하거나 평가를 내린 아이템이 적은 경우, 혹은 아이템의 종류가 너무 많은 경우에는 제대로 된 추천 결과를 내지 못하는 이른바 ‘회박성 문제(sparsity problem)’ 역시 기존 협동 필터링 접근법의 중대한 한계라고 할 수 있다(Balabanovic and Shoham, 1997; Kim and Yum, 2005; Adomavicius and Tuzhilin, 2005).

위의 단점들 외에도 기존의 두 방법론들은 다음의 2가지 공통적인 한계점을 지니고 있다. 우선 두 알고리즘 모두 최소한 한 건 이상의 구매를 한 고객에 대해서만 추천이 가능하다. 왜냐하면, 내용 기반 접근법의 경우는 유사한 아이템을 찾기 위해, 협동 필터링의 경우는 유사한 고객을 찾기 위해 추천을 위한 ‘최소한의 구매 이력’을 필요로 하기 때문이다. 또한, 많은 쇼핑몰들이 회원가입 단계를 통해, 고객으로부터 기본적인 인적사항 및 기타 개인특징정보를 입력받고 있음에도 불구하고, 상기 두 방법론은 이러한 정보를 추천에 거의 사용하지 않고 있다. 이러한 한계점은 특히 구매 비율이 저조하고, 고객층이 넓지 않은 국내의 중·소형 인터넷 쇼핑몰들에게는 치명적이라고 할 수 있다.

이러한 상황에서 본 연구는 위에 제시된 기존 추천 기법들의 공통적인 한계점을 어느 정도 극복하면서도, 효과적인 추천 결과를 생성할 수 있는 새로운 추천 시스템을 제안하고자 한다. 아울러 실제적인 적용 가능성에 역점을 두고 고안된 본 추천 시스템의 효용성을 검증하기 위해, 국내 한 전문 쇼핑몰에 본 연구에서 제안하는 모형을 실제로 적용해 보고, 이를 기반으로 만들어진 프로토타입(prototype) 시스템에 대한 온라인 설문을 실시해, 과연 본 연구에서 제안하고 있는 새로운 추천 시스템이 사용자들에게 가치있는 유용한 정보를 제공하는지 검증해 보고자 하였다.

본 연구에서 적용할 사례는 국내 대표적인 디이어트 전문 온라인 쇼핑몰 중 하나인 G 온라인 쇼핑몰의 사례이다. G 온라인 쇼핑몰은 디이어

트 전문 온라인 쇼핑몰 중에서는 최대 규모이지만 일반적인 온라인 쇼핑몰과의 규모로는 중형에 속하는 쇼핑몰로서 협동 필터링이나 내용기반필터링에서의 한계점이 될 수 있는 여러 가지 요소들을 포함하고 있다. 특히, 본 대상사례의 경우, 상품의 구매 비율은 저조하지만, 다이어트라는 주제로 맞춤 정보 서비스 등을 제공하기 때문에, 고객의 개인적인 프로필과 관련한 풍부하면서도 양질의 정보를 보유하고 있는 상황이다. 때문에, G 쇼핑몰에서는 이러한 고객정보를 상거래와 추천시스템을 통해 자연스럽게 연계하여 고객의 만족을 높이는 동시에, 매출도 향상시키는데 많은 관심과 노력을 기울이고 있다. 이런 취지에서, 본 연구에서는 G 온라인 쇼핑몰에 본 연구에서 제안하는 추천 시스템을 적용하여 중소형 온라인 쇼핑몰에서의 응용 가능성을 살펴 볼 것이다.

본 연구의 구성은 다음과 같다. 우선 I장에서는 본 연구의 소개를 하였고, II장에서는 추천시스템 분야의 각종 문헌 고찰을 통해, 기존 접근법의 한계점과 함께 본 연구에서 제시하고자 하는 모형과 관련한 기본적인 개념들을 함께 살펴 볼 것이다. 이어 III장에서는 본 연구에서 제안하는 새로운 상품추천시스템의 모형이 제시되며, IV장에서는 본 연구에서 제안한 모형을 실제 사례에 적용해 보고, 그 결과를 살펴보고자 한다. 끝으로 V장에서는 본 연구의 의의와 한계점을 고찰해 보고자 한다.

## II. 문헌고찰

### 2.1 전통적인 추천기법과 한계점

인터넷 쇼핑몰에서의 상품추천시스템은 실생활에서도 활발하게 이용되고 있는데, 그 대표적인 예로는 아마존(Amazon), 무비렌즈(MovieLens), 시디나우(CD Now), 제이씨 페니(JC Penny) 등이 있다. 상품추천시스템의 가장 핵심적인 부분

은 추천알고리즘이라 할 수 있으며, 대표적인 추천기법으로는 내용기반 필터링과 협업필터링 등의 방법이 있다. 이 중, 내용기반 필터링은 과거에 대상 고객이 선호했던 아이템과 가장 유사한 아이템을 찾아서, 추천하는 방식이다. 즉, 내용 기반 기법은 상품과 상품 사이(item-to-item)의 관계를 토대로 추천 결과를 생성하는 방식인데, 이러한 내용 기반 기법에서는 그 무엇보다도 사용자(고객)의 선호도와 아이템(상품)의 특성을 어떻게 모델링 할 것인가가 이 기법의 성과를 결정짓는 핵심 요인이라 할 만큼 매우 중요하다.

이러한 내용 기반 추천 기법이 갖는 가장 큰 장점은 상품 자체를 모델링하는 기법이기 때문에, 직접적이고 단순하다는 점이다(Wu *et al.*, 2001). 하지만, 태생적으로 지니는 중요한 한계점들도 있다. 우선, 내용 기반의 추천 방식은 추천을 위한 분석의 깊이가 얕을 수 밖에 없다(shallow analysis)는 한계점이 있다. 왜냐하면 앞서 지적했듯이, 내용 기반 추천 알고리즘을 적용하기 위해서는 추천의 대상이 되는 상품이나 아이템에 대한 특성을 추출해야 하는데, 이것이 사실상 효과적으로 이루어지기 어렵기 때문이다. 또 다른 내용 기반 접근법의 한계점으로는 내용 기반 접근법의 추천 결과가 너무 과도하게 특정 부분에 치우치게 된다는 점(overspecialization)을 지적할 수 있다. 이는 추천의 원리가 고객이 이전에 좋게 평가한 상품과 비슷한 상품군을 찾는 방식으로 이루어지기 때문에, 예전에 어떤 상품을 평가했었는가 하는 것에 추천 결과가 너무 의존하게 된다는 내용 기반 추천 방식의 한계점을 의미한다. 이러한 내용기반 필터링 기법의 한계점들로 인해, 현실적으로는 협업필터링이 선행연구에서 더 활발하게 이용되고 있다(김재경 등, 2002; 김종우 등, 2004; 조윤호 등, 2004; 김재경 등, 2005 참고).

협업필터링은 사용자 사이의 연관성을 기반으로, 선호도 또는 구매 패턴이 유사한 고객군을

분류하고, 유사한 고객에 속하는 다른 사람들이 선호하는 상품을 추천하는 방법이다(Funakoshi and Ohguro, 2000). 협업필터링에 대한 초기 연구로는 Tapestry(Goldberg *et al.*, 1992), GroupLens (Resnick *et al.*, 1994) 등의 사례가 대표적이며, Ringo와 Video Recommender 등과 같은 이메일과 웹 기반의 협업필터링 기법에 의한 추천시스템 등이 있다(Sarwar *et al.*, 2000). 협업필터링은 일반적으로 고객들이 동질적인 평가결과를 보이는 상품군에 대해 상대적으로 높은 예측력을 보이며, 데이터가 충분한 경우에는 다른 기법에 비해 상대적으로 높은 예측력을 보이는 장점을 가지고 있다(Konstan *et al.*, 1997; Pazzani, 1999). 이에 따라 협업필터링은 상품추천시스템 관련 연구에서 활발하게 이용되고 있으나 아래와 같은 한계점도 가지고 있다.

협업필터링은 기본적으로 상품에 대한 고객의 선호도 또는 구매이력자료를 바탕으로 추천을 하게 되므로 구매이력을 많이 보유하고 있는 대형 인터넷 쇼핑몰에서는 유용하지만, 구매이력이 상대적으로 부족한 중소 인터넷 쇼핑몰이나 사업 초기단계의 인터넷 쇼핑몰의 경우에서는 적용 가능성성이 떨어진다고 할 수 있다. 즉, 협업필터링의 속성상 구매이력이 부족한 경우에는 추천의 성과가 떨어질 수 밖에 없으며 이런 점은 이미 선행연구에 의해 협업필터링의 가장 중요한 문제점 중 하나로 지적되고 있다(김재경 등, 2002; Cho *et al.*, 2002; Kim *et al.*, 2002; 김재경 등, 2003; 김종우 등, 2004; 조윤호 등, 2004; Cho and Kim, 2004; 김재경 등, 2005; Kim and Yum, 2005; Adomavicius and Tuzhilin, 2005 참고). 이런 문제점을 흔히 희박성(sparsity) 문제라고 하며, 희박성 문제를 완화하기 위해 선행연구에서는 웹 로그 정보를 활용하여 간접적으로 선호도 데이터를 보충하고자 하였다(Cho *et al.*, 2002; Kim *et al.*, 2002; 김재경 등, 2003; Cho and Kim, 2004; 김재경 등, 2005). 그러나 웹 로그 정보는 일반적으로 대용량이며 정

제되지 않은 형태이어서 고객의 선호도 점수를 직접 취득하는 것만큼 전처리에 많은 시간과 비용이 소요된다는 단점이 있다.

협업필터링의 중요한 한계점 중 다른 하나는 고객과 거래 데이터가 증가함에 따라 유사한 고객군을 찾기 위한 연산량이 기하급수적으로 증가하는 현상이 발생할 수 있다는 것이다(김재경 등, 2002; Cho *et al.*, 2002; Kim *et al.*, 2002; 김재경 등, 2003; 조윤호 등, 2004; Cho and Kim, 2004; 김재경 등, 2005; Adomavicius and Tuzhilin, 2005 참고). 선행연구에서는 이런 문제점을 확장성(scability) 문제라 하였는데, 이는 해결해야 하는 문제가 제시된 이후에야 추론을 시작하는 ‘게으른 학습방법(lazy learning technique)’의 일반적인 특징으로, 신속한 응답을 원하는 인터넷 사용자의 특성을 감안할 때 고객의 이탈을 유도할 수 있는 치명적인 한계점이다. 선행연구에서는 이 문제점을 보완하기 위해서 아래와 같은 여러 방법들을 제안하였다.

김재경 등(2002), Cho 등(2002), Kim 등(2002), 조윤호 등(2004), Cho와 Kim(2004), 김재경 등(2005) 등은 확장성과 희박성의 문제를 보완하기 위하여 상품계층도(product taxonomy)를 활용하는 방법을 제안하였으나, 여전히 하나의 상품계층도의 각 상품계층군 안에서는 각 고객의 선호도가 제대로 반영되지 않아서 추천의 성과가 떨어지는 경우가 발생할 수 있다. 또, 상품계층도의 작성이 상품추천의 성과에 큰 영향을 미칠 수 있는데 선행연구에서는 전문가의 주관적인 판단을 참고하는 방식으로 연구를 진행하였으나 이 점 역시 한계점이 될 수 있다.

확장성 문제를 완화하기 위한 또 다른 노력으로는 군집분석을 협업필터링의 사전과정으로 수행하여 탐색공간을 줄이는 방법이 있다. 김재경 등(2003)은 K-Means 군집분석을 협업필터링 사전단계로 활용하여 탐색공간을 축소하였고, Roh 등(2003)과 강부식(2003)은 군집분석기법의 하나인 자기조직화지도를 활용하여 사례탐색공

간을 축소하였다. Kim과 Han(2001)은 협업필터링과 함께 ‘계으른 학습방법’의 하나인 사례기반추론에서 자기조직화지도 분석을 추론 이전 단계에 활용하여 분석 데이터의 양을 줄이고자 하였다. 그러나 이 같은 선행연구들은 ‘계으른 학습방법’에서 확장성 문제의 심각성을 인지하고 이에 대한 보완방법으로 군집분석의 방법을 이용하는 것이 유용하다는 주장만 하였을 뿐, 제안되었던 상품추천시스템의 성능을 향상시키는데 결정적인 요인이 될 수 있는 군집분석의 방법론적 개선에는 큰 관심을 두지 않았다.

선행연구들의 결과를 종합해 보면 협업필터링이 상품추천에 있어서 주요한 방법이지만, 회박성과 확장성의 문제가 한계점이며, 이 한계점을 보다 정교하게 보완하여 회박성과 확장성의 정도를 완화시키고 이에 따라 추천의 성능을 제고할 수 있는 방안에 대한 연구가 필요하다고 할 수 있다.

## 2.2 데이터 마이닝을 이용한 추천 시스템 – 연관규칙 기법

데이터 마이닝(data mining)은 자동화된, 혹은 반자동화된 수단을 통해 방대한 양의 데이터 속에서 의미 있는 패턴이나 규칙을 찾아내기 위해 수행하는 일련의 탐색 및 분석 과정을 의미한다. 이른바 지식 탐색(knowledge discovery)이라고도 불리우는 데이터 마이닝은 오늘날 인터넷으로 인해 양질의 고객 데이터가 방대한 규모로 축적되고 있는 상황에서 특히 관심이 모아지고 있다. 데이터 마이닝 기법으로는 기본적인 통계적 기법부터 시작해서 인공지능 기법까지 다양한 종류들이 있는데, 이러한 기법들을 통해 우리는 분류(classification), 추정(estimation), 예측(prediction), 분류(clustering) 등 여러 가지 목적을 달성할 수 있다(Berry and Linoff, 1997).

인터넷 상거래 사이트에서 고객에게 맞춤 서비스를 제공하는 목적으로도 데이터 마이닝 기

법은 매우 효과적으로 사용될 수 있는데, 그 중에서도 특히 군집화(clustering) 기법, 연관 규칙(association rules) 기법, 분류(classification) 기법 등은 가장 대표적으로 활용되는 기법이라고 할 수 있다. 여기서 군집화 기법의 경우는 주로 인터넷 쇼핑몰 고객들의 유사성 검색, 고객 세분화, 패턴 인식, 추세 분석 등에 활용되며, 연관규칙기법은 장바구니 분석(market basket analysis)이나, 고객 프로필과 구매 행동 간의 연관성 검색 등에 활용된다. 이에 비해, 분류 기법은 주로 고객의 구매 행동 예측에 활용되며, 통계적 회귀(regression) 모형이나 의사결정나무(decision tree), K-nearest neighbor 기법 등이 일반적으로 적용된다(Yu, 1999).

이러한 기법들 중에서, 연관규칙기법은 일반적으로 구매 행위에 있어서 특정 아이템과 다른 아이템 간에 어떤 연관 관계가 있는지를 찾아보는 것이다. 연관규칙기법은 비록 다른 데이터 마이닝 기법에 비해 단순하지만, 적용이 쉽고, 상당히 의미 있는 정보를 제공해 준다는 측면에서 인터넷 쇼핑몰의 상품 추천 시스템에서 가장 널리 적용되고 있는 기법 중 하나이다(Wang et al., 2004; Liu and Shih, 2005).

연관규칙기법의 알고리즘은 다음과 같다.

$$I = \{i_1, i_2, i_3, \dots, i_m\}, T \subseteq I \quad (1)$$

위 식 (1)의  $I$ 는 전체 상품 아이템들의 집합이고, 이 때 각각의 구매 거래를 나타내는  $T$ 는  $I$ 의 부분 집합이 된다. 이런 상황에서 연관규칙은 아래와 같은 형태를 취하게 된다.

$$X \Rightarrow Y \text{ where } X \subseteq I \text{ and } Y \subseteq I \quad (2)$$

즉, 위 식 (2)와 같은 연관규칙이 존재할 경우, 우리는 구매 거래 중에 일련의 아이템이 고객에 의해 선택되었다고 할 때, 어떤 다른 아이템들이 함께 선택될 수 있을 것인지를 체계적으로 판단할 수 있는 것이다(Yu, 1999).

일반적으로 연관규칙 분석의 결과를 판단할

〈표 1〉 연관규칙 분석 결과의 판단 기준

지표	설명
지지도 (Support)	연관규칙 $A \Rightarrow B$ 의 지지도는 고객이 $A$ 와 $B$ 아이템을 동시에 갖고 있을 확률로 정의 ( $= \Pr(A \cap B)$ )
신뢰도 (Confidence)	연관규칙 $A \Rightarrow B$ 의 신뢰도는 고객이 $A$ 를 가지고 있을 때, $B$ 도 가지고 있을 확률로 정의 ( $= \Pr(A   B)$ )
향상도 (Lift)	연관 규칙의 강도(strength)를 측정하는 지표. 예를 들어, 규칙 $A \Rightarrow B$ 에 대한 Lift 값이 2라고 한다면, $A$ 를 가진 고객이 $B$ 를 가지고 있을 확률이 무작위로 선택된 고객에 비해 2배만큼 높다는 것을 의미 ( $= \frac{\Pr(B   A)}{\Pr(B)}$ ).

때, 참고하는 지표로는 세가지가 있는데, 이 지표들은 다음의 〈표 1〉과 같다.

일반적으로 연관규칙 마이닝(association rule mining)은 크게 두 단계에 걸쳐 진행되는데, 우선 첫 단계에서는 사용자가 정의한 최소 지지도 값을 넘어선 규칙들만 골라내는 작업이 이루어진다. 그렇게 해서 대표성을 지니는 규칙들이 찾아지면, 그 다음에는 그 중에서 사용자가 정의한 최소 신뢰도 값을 넘어선 규칙들을 골라내는 2차 선별 작업이 이루어 지게 되고, 이 작업 까지 끝나면 연관 규칙 마이닝이 최종적으로 끝나게 된다. 이 때 단계 2에서는 신뢰도 값 대신 향상도 값을 기준으로 활용할 수 있는데, 특히 연관 규칙에 연계되어 있는 아이템들을 많은 고객들이 보유하고 있는 상황인 경우, 향상도 값을 기준으로 활용하는 것이 더 의미 있는 규칙을 추출하는데 유리하다(Wielenga *et al.*, 1999).

### 2.3 고객의 프로필 정보를 활용한 추천 기법

고객의 프로필 정보를 기반으로 추천 결과를 생성하는 추천 시스템의 경우, 실무에서는 오래 전부터 다양하게 활용되어 왔지만, 학계에서는 그렇게 다양한 연구가 진행되지 않았다. 대표적인 프로필 기반 추천 시스템과 관련된 선행 연구로는 크게 다음의 두 가지를 살펴볼 수 있다.

우선 Krulwich(1997)는 미국 전역의 인구를

대상으로 약 600개 이상의 변수를 포함한 4만여 건의 인구 프로필 데이터 베이스를 토대로, 이를 총 62가지 종류의 인구통계 군집으로 분류하고, 이 군집을 기준으로 새로운 고객의 프로필을 결정해 이를 추천에 활용하는 새로운 추천 알고리즘을 제안하였다. 그는 이러한 프로필 기반 추천 알고리즘을 ‘Lifestyle Finder’라는 웹 사이트 추천 엔진에 적용해, 전세계 약 2만여 명의 방문객들을 대상으로 그 유용성을 평가해보았는데, 실험 결과 무작위로 선정된 URL보다 추천 알고리즘을 통해 선정된 URL이 보다 높은 선호도와 클릭율을 보이는 것으로 조사되었다.

한편 Aggarwal 등(1998)은 기존의 연관규칙 분석 알고리즘을 개선해, 프로필 정보와 구매 행위 사이의 연관관계를 규칙으로 유도할 수 있는 프로필 기반의 연관규칙 알고리즘(profile-based association rule)을 제안하였다. 즉, 기존의 연관규칙 분석이 아이템과 아이템 간의 관계를 살펴보는 방식(item set approach) 이었던 것에 비해, 여기서는 고객 프로필과 관련된 범주형 변수나 정량적인 변수도 함께 고려하는 방식을 도입했다는 점에서 기존 방식과 차별화된다고 할 수 있다. 또한 이 연구에서는 프로필 기반 연관규칙을 찾아내는데 있어서 나무구조 형태의 다차원적도구조(multidimensional index structure)를 도입, OLAP과 같은 온라인 처리 환경하에서 보다 효과적으로 적용될 수 있는 알고리즘을 제

안했다는 의의도 아울러 찾아 볼 수 있다.

그 외, Stegmann 등(2003)은 고객 프로필 정보를 제품정보와 함께 활용하여, 구매 단계별로 차별화된 추천이 가능한 시스템 구조를 제안하였으며, Kim 등(2004) 역시 모바일 환경에서 고객의 프로필 정보를 이용해 모바일 컨텐츠를 제공할 수 있는 새로운 통합 추천시스템을 제안하는 등 고객 프로필 정보만을 활용하지는 않았지만, 기타 다른 정보와 고객 프로필 정보를 결합하여 추천시스템에 적용한 사례는 다수 찾아볼 수 있다.

### III. 새로운 상품추천시스템의 프레임워크

이상 살펴본 데이터마이닝 기법과 프로필 정보를 기반으로 한 추천시스템들에 대한 기존 연구를 토대로, 본 연구에서는 앞서 연구의 목적에서 설명했던 기존 접근 시스템의 한계점들을 극복할 수 있는 새로운 시스템을 제안하고자 한다. 새로운 상품추천시스템의 전체적인 구조를 소개하기에 앞서, 본 연구를 통해 제시하고자 하는 상품추천시스템이 달성하고자 하는 4가지 핵심 목표와 이를 위해 적용한 기법들을 간략하게 소개하면 다음과 같다.

- 구매이력정보가 부족하더라도 효과적인 상품 추천이 가능하도록 고객 프로필 관련정보를 최대한 활용한다. 본 연구에서 제시하는 모형은 기존의 구매 이력을 활용하는 추천기법 외에, 이를 보완할 수 있는 '프로필 기반 추천 알고리즘'을 도입해, 인터넷 쇼핑몰 업체에서 보유하고 있는 고객정보를 최대한 활용 할 수 있는 모델로 개발한다. 이는 Krulwich (1997), Aggarwal 등(1998), Stegmann 등(2003), Kim 등(2004) 등에서 제안된 방법을 참고한 것이다.
- 역사가 길지 않고, 영세한 인터넷 쇼핑몰에서

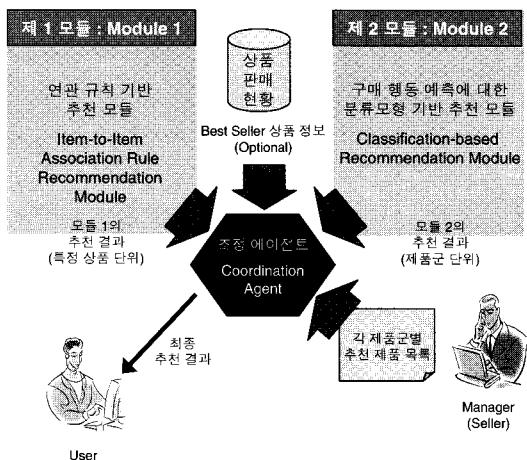
도 효과적으로 활용할 수 있도록, 신규 고객이나 구매 이력이 없는 고객에게 효과적인 제품 추천을 할 수 있는 추천 알고리즘을 구축한다. 본 연구에서 제시하는 모델은 궁극적으로 '광범위한 추천'을 지향하는 모델로 개발하여, 구매 경험이 전혀 없는 고객이 사이트에 방문했을 경우에도, 그에게 유익한 추천 정보를 제시할 수 있도록 설계한다. 이는 프로필 기반의 추천 알고리즘을 도입할 경우, 달성 가능한 목표이다(Schafer *et al.*, 2001; Aggarwal *et al.*, 1998; Krulwich, 1997).

- 최대한 간단하게 알고리즘을 구현해, 적은 연산량으로도 추천 결과를 생성할 수 있도록 설계한다. 본 연구에서 제시하는 모델은 '모델 기반의 추천 알고리즘'을 지향했던 Iyengar와 Zhang(2001)의 연구와 마찬가지로 실제 적용하고자 할 때, 단순히 입력 변수의 값만 입력되면, 간단한 일련의 연산들만으로도 추천 결과가 도출될 수 있도록 효율적인 모델로 개발하도록 한다.
- 추천 방식이나 절차에 유연성을 부여해, 고객에게도 유용하지만, 판매자의 의도도 일부 반영할 수 있는 추천 모델을 설계하도록 한다. 본 연구에서 제시하고자 하는 모델은 마치 기존의 Konstan 등(1997)의 GroupLens 연구에서 차원을 줄이기 위해 아이템을 일정 기준의 그룹으로 묶었던 것처럼, '상품군(product group)'의 개념을 도입하도록 한다. 그래서, 구입 예상 상품군을 구체적인 추천 상품의 형태로 변환하는 과정에서 판매자의 의도가 일부 반영될 수 있도록 설계한다.

이상의 목표를 달성할 수 있도록, 새롭게 설계된 본 연구의 상품 추천 모델의 구조는 다음의 <그림 1>과 같이 구성된다.

<그림 1>의 구조도에서 볼 수 있는 것과 같이, 본 연구에서 제시하고자 하는 새로운 모델은 두 가지 종류의 서로 다른 데이터 마이닝 알

고리즘을 통해 구현되는 추천 모듈들과 이들의 결과를 조정해서 궁극적으로 최종적인 추천 결과를 양산해 내는 조정 에이전트로 되어 있다. 이상 세 요소들의 기능과 요소들 사이의 관계를 보다 구체적으로 살펴보면 다음과 같다.



〈그림 1〉 새로운 상품 추천 모델의 구조도

### 3.1 모듈 1: 연관규칙 기반 추천

모듈 1은 아이템 간의 연관규칙을 찾아, 그 결과로 추천 아이템을 생성하는 모듈이다. 즉, 전체 구매 데이터를 기준으로 구매된 아이템들 간에 어떤 연관규칙들이 있는지를 '모델화 했다가, 대상 고객이 방문했을 때 그 고객이 이전에 어떤 상품을 구매했는지를 확인해, 그 상품이 속해있는 연관관계들을 토대로 새로운 상품을 추천하는 모듈인 것이다.

이러한 연관규칙 기반의 추천은 최소한의 구매 정보만 가지고도 의미 있는 지식을 추출할 수 있고, 결과가 적용하기 쉬운 규칙의 형태로 생성된다는 장점으로 인해, 기존의 많은 추천 시스템 연구에서 적용된 바 있다. 하지만 기존의 연구들이 협동 필터링과 연관 규칙 분석을 접목한다든가(Lee *et al.*, 2001), 군집분석기법과 연관규칙 분석을 결합하는 등(Lawrence *et al.*,

2001), 주로 변형된 형태의 연관규칙기법을 제안하는 것에 초점을 맞추고 있다면, 본 연구에서는 일반적인 형태의 아이템과 아이템 간의 연관규칙을 적용하되, 이 결과를 조정 에이전트에서 다른 추천 모듈의 결과와 함께 접목시켜 추천을 하도록 설계했다는 측면에서 기존 연구와 차별화 된다고 할 수 있다.

모듈 1은 이처럼 아이템과 아이템간의 연관규칙을 기반으로 추천을 하는 방식이기 때문에, 추천의 결과가 '직접적인 아이템 형태'로 이루어지게 된다는 특징이 있다. 또한 이는 아이템을 중심으로 모델링한 추천 알고리즘이라는 측면에서 내용 기반 필터링 방식과 유사한 면이 있다.

### 3.2 모듈 2: 구매 행동 예측을 위한 분류 기반 추천

모듈 1은 비록 추천 결과가 직접적인 형태로 도출되고, 적용이 용이한 규칙의 형태로 모델이 표현된다는 장점이 있지만, 고객이 최소한 어느 정도의 구매 이력을 갖고 있는 경우에만 추천 결과를 제시할 수 있다는 단점이 있다. 그런데 이러한 단점은 앞서 본 연구의 배경에서 지적했듯이, 구매 비율이 저조하고, 고객층이 넓지 않은 국내의 중·소형 인터넷 쇼핑몰들에게는 매우 치명적인 한계점이라 아니 할 수 없다. 그래서 본 연구에서는 구매 이력이 충분치 않아도 고객의 상세한 프로필 정보만 있으면 양질의 추천 결과를 제시할 수 두번째 모듈을 제안한다.

모듈 2는 궁극적으로 '사용자의 프로필' 정보를 토대로 모델링된 고객군들의 구매 추세를 기반으로 추천하는 알고리즘으로, 이는 사용자를 중심으로 추천 알고리즘을 모델링 하는 협동 필터링의 개념이 일부 반영된다. 즉, 추천 여부를 구매는 1, 비구매는 0으로 분류하는 모형의 결과를 이용하는 것이다. 모듈 2와 관련하여, 또

하나 본 연구에서 제안하고자 하는 것은 예측의 대상을 구체적인 상품 자체가 아닌 ‘상품군(product group)’으로 하자는 것이다. 이렇게 예측의 대상을 ‘상품군’으로 하는 것은 크게 2가지 의미가 있다. 우선 개별의 상품이 아닌 ‘상품군’을 예측하게 되면, 차원 감소(dimension reduction) 효과를 가져와, 모델링에 활용할 수 있는 표본의 수가 늘어나므로 보다 일반화된 모델을 구축할 수 있다. 실제로 이처럼 표본의 수를 늘리고, 데이터의 밀도를 높이는 효과를 얻기 위해 추천의 대상을 그룹화해서 분석하는 경우는 기존 연구에서도 쉽게 찾아볼 수 있는데, 대표적으로 ‘협동 필터링’ 연구의 대표격인 GroupLens 연구를 들 수 있다(Konstan et al., 1997).

또한 ‘상품군’을 기준으로 예측모델을 만들게 되면, 추천 결과를 ‘구체적인 개별 상품’으로 해석하는 과정에서 판매자의 의도를 반영할 수 있다는 장점도 있다. 즉, 기존 추천알고리즘의 한 계점이라 할 수 있는 추천의 결과에 판매자의 의도가 제대로 반영되지 못한다는 점을, 본 연구에서 제안하는 것과 같이 상품군의 개념을 도입해 추천 시스템을 설계하면, 특정 상품군을 고객이 구매할 것으로 예상했을 때 그 상품군에 속하는 상품들 중에서 판매자가 자신의 의도를 반영한 상품을 선별해 추천할 수 있는 것이다. 이처럼 상품군 개념을 이용하면 판매자의 의도를 반영할 수 있는 장점은 있으나, ‘간접적인 추천 결과’를 만들게 됨으로서, 앞서 소개한 모듈 1의 결과 보다 추천 정보의 정교성이 떨어질 수 있다는 단점이 발생할 있다. 본 연구에서는 이러한 단점을 보완하기 위해 뒤에 설명하고 있는 ‘조정 에이전트’라는 개념을 도입해 활용한다.

모듈 2를 위한 분류모델을 구축하고자 할 때에는, 구매/비구매의 이진값을 예측하는 모형이 필요한데, 이러한 역할을 할 수 있는 모형들로는 로지스틱 회귀분석을 비롯해, 인공신경망, 의사결정나무 등 다양한 인공지능기법들이 존재

한다. 따라서 가장 효과적인 예측모델을 구축하기 위해서는, 이 기법들을 다양하게 적용해 본 뒤 분류성과가 가장 우수한 기법을 최종적으로 선택해 분류 모델로 활용하는 것이 바람직하다.

### 3.3 조정 에이전트(Coordination Agent)

지금까지 소개한 모듈들이 실질적인 추천 결과를 양산해 내는 역할을 하는 것에 반해, 조정 에이전트는 마치 Balabanovic와 Shoham(1997)의 Fab 시스템에 대한 연구에서 제안된 ‘선택 에이전트(selection agent)’의 개념처럼, 각 모듈들이 추천한 결과들을 한데 모아, 그 중 우수한 것을 고르고, 최종적으로 고객에게 보여 줄 추천 결과를 정리하는 기능을 수행하게 된다. 좀 더 세부적으로 조정 에이전트의 역할을 정리해 보면 다음과 같다.

- 예측된 구매 예상 상품군을 구체적인 추천상품 리스트로 전환  
모듈 2는 ‘상품군’ 단위로 추천 결과를 생성하므로, 이를 구체적인 상품 아이템 단위의 추천 결과로 전이하는 과정이 필요하다. 이 과정을 조정 에이전트가 담당하게 되며, 이 때 조정 에이전트는 미리 판매자(관리자)가 입력한 ‘각 상품군별 제품추천 리스트’를 참고하여 추천을 하게 된다.
- 두 모듈에서 아무런 추천 결과도 생성하지 못한 경우, 가장 많이 팔린 제품들을 중심으로 추천 결과 생성  
모듈 1은 반드시 구매 이력이 있어야만 추천 결과를 생성할 수 있기 때문에 상황에 따라 아무런 추천 결과를 만들어 내지 않을 가능성이 있으며, 모듈 2 역시 고객의 프로필을 기반으로 각 상품군의 구매 여부를 개별적으로 예측하므로 아무런 추천 결과를 만들어 내지 않을 가능성이 있다. 이처럼 두 추천모듈에서 아무런 추천 결과를 생성하지 못할

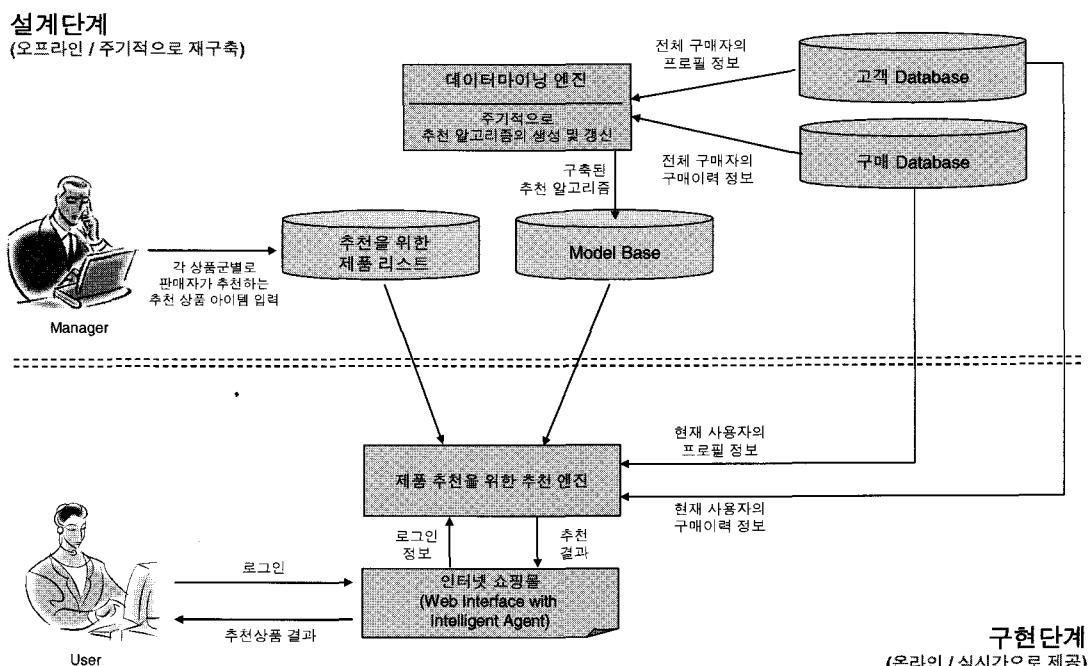
경우, 조정 에이전트는 ‘사용자 집단 전체의 추세(community-wide trend)’를 반영해 추천 리스트를 결정한다는 측면에서 베스트셀러 중심의 추천 결과를 생성하게 된다.

- 우선 순위에 의해, 각 모듈에서 생성된 추천 결과들을 조정하여 최종 추천 리스트를 결정 각 모듈들에 의해 생성된 추천 결과가 모두 ‘조정 에이전트’ 쪽으로 취합되게 되면, 조정 에이전트는 추천의 우선 순위에 따라 고객에게 그 중 어떤 것을 골라서, 어떤 형태로, 그리고 어떤 순서로 보여줄 것인지를 결정하게 된다. 본 연구에서 제안하고자 하는 모델에서 는 직접적인 추천 결과를 생성함으로서 상대적으로 더 정확한 추천 결과를 제시할 수 있는 모듈 1의 추천 결과를 1순위로 보여주고, 간접적이지만 상품군 별로 추천 결과를 생성해 주는 모듈 2의 추천 결과를 2순위로 보여주도록 제안하고 있다. 그리고, 이 때 구매할

것으로 예상되는 상품군이 여러 개로 나타날 경우에는, 추천 결과에 다양성을 반영하기 위해, 각 상품군에서 동등한 개수의 추천 결과가 포함될 수 있도록 조정한다. 마지막으로 두 추천 모듈들이 아무런 추천 결과를 생성하지 못할 경우에는 조정 에이전트가 직접 고객의 구매 현황 데이터베이스에 접근하여 전체적으로 가장 많이 팔린 제품들을 중심으로 생성된 추천 리스트를 추천하게 된다.

지금까지 살펴본 본 연구의 상품 추천 모델은 전체적인 웹 시스템으로 구축될 때, 다음의 <그림 2>와 같은 구조로 구현된다.

아래의 구조는 Schafer 등(2001)가 제시한 전형적인 추천시스템의 구조를 충실히 따르고 있다. 다만 전통적인 추천시스템의 구조에서는 고객과의 접점 역할을 하는 인터페이스 기능이 단순히 추천결과를 보여주는 기능만을 수행하는 반면, 본 연구에서 제안하는 시스템에



<그림 2> 새로운 상품추천시스템의 구조

서는 추천을 담당하는 모듈이 2개이므로, 인터페이스가 추천 결과를 조정하고, 화면 출력의 체계를 스스로 결정짓는 ‘조정 에이전트’의 개념을 함께 포함하고 있다는 차이가 있다고 할 수 있다.

#### IV. 사례 적용

본 장에서는 제안하는 추천시스템의 활용 가능성을 검증하기 위해 앞서 제시한 추천 모델을 실제 사례에 적용해 보고자 한다. 본 연구의 모델을 적용할 사례는 초기에 제시했던 연구의 목적에 부합될 수 있도록 상대적으로 고객의 구매 건수가 많지 않은 국내 전문 중소형 인터넷 쇼핑몰 중 하나를 선정하고자 하였다. 그리하여, 상대적으로 양질의 다양한 고객 정보를 확보하고 있는 G 인터넷 쇼핑몰 사이트를 본 연구의 대상 사례로 최종 선정하였다. 그런 다음 연구 대상의 사이트로부터 특정 기간의 고객 정보와 구매 정보를 제공받아, 이를 토대로 본 연구의 추천 모델에서 제안한 두 가지 추천 모듈을 위한 추천 모델을 상용화된 통계 및 데이터 마이닝 패키지를 이용해 구축하였다. 이렇게 구축된 모델은 웹 기반의 프로토타입으로 제작되어 실제 사용자가 활용하도록 하였으며, 프로토타입에 간단한 온라인 설문 기능을 첨부하여 본 연구에서 제안한 추천 시스템의 효용성을 사용자들이 어떻게 평가하는지를 살펴보고자 하였다.

##### 4.1 사례의 개요

본 연구의 대상 사례인 G 온라인 쇼핑몰은 지난 2000년 7월에 정식 오픈한 국내 유명 다이어트 전문 온라인 쇼핑몰이다. 오프라인에서 이미 다이어트 분야에 확고한 입지를 가지고 있는 P사가 운영하고 있는 G 사이트는 방대한 컨텐츠와 우수한 서비스, 그리고 운영사의 신뢰도 높은 브랜드 이미지로 인해 국내 온라인

다이어트 분야를 선도하고 있는 전문 인터넷 쇼핑몰이다. 다이어트는 그 특성상 고객이 익명성을 선호하면서도 한편으로는 성공적인 다이어트를 위해 서비스제공자와의 상호작용 및 의사소통을 많이 원한다는 측면에서, 인터넷(웹)이 매우 효과적인 매체로 활용될 수 있는 분야이다. 특히, 이러한 전문 쇼핑몰의 경우, 회원가입시 고객이 제대로 된 맞춤 서비스를 받기 위해 양질의 고객정보를 풍부하게 서비스 제공자에게 제공해 주기 때문에, 이러한 고객 정보는 서비스 제공자에 있어 매우 유익한 정보로 활용될 수 있다.

본 연구의 모델링을 위해 G 사이트로부터 확보한 데이터는 최근 약 3개월간 구매한 총 3298 명의 회원의 4353건에 대한 구매 내역 데이터였다. 이 때, 데이터의 필드는 전처리 이후를 기준으로 해서 총 46개였는데, 이 중 41개의 변수가 명목형 변수였고, 나머지 5개가 비율척도로 된 변수였다. 데이터는 상품과 관련된 정보를 비롯해, 구매자의 연령, 체중, 키, 질병유무, 원하는 감량 목표, 다이어트 요법 경험 유무 등 고객의 개인적인 특성과 관련된 정보가 대부분을 이루고 있다.

현재 분석 대상이 되고 있는 G 사이트의 경우, 전체 구매 고객의 80.44%가 위 기간동안 단 1건만 구매경험이 있는데다, 그나마 구매 경험이 1건이라도 있는 고객의 비중은 전체 회원의 단 5%수준에 불과한 상태인 것으로 나타났다. 이는 앞서 연구의 목적에서 제시했던 기존 추천 방식들의 ‘최소한 1건 이상의 거래 내역이 있어야만 추천이 가능하다’는 특성이 실제로 상당히 치명적인 문제점임을 증명하고 있다고 할 수 있다.

##### 4.2 추천 모델의 구축

###### 4.2.1 연관규칙 마이닝을 활용한 규칙 생성

우선 추천 모델의 모듈 1에 적용되는 연관규칙 분석은 SAS E-Miner 4.0 패키지를 적용해 도출하였다. 이 때, 규칙 선정을 위한 기준은 지지도를

0.5%, 신뢰도를 10%로 설정하였는데, 그 결과 <표 2>와 같이 총 14개의 규칙을 도출할 수 있었다.

#### 4.2.2 구매행동 예측을 위한 분류

일반적으로 구매행동 예측을 위한 분류를 위해 적용될 수 있는 알고리즘으로는 다양한 방법들이 있지만, 본 연구에서는 로지스틱 회귀분석과 인공신경망, 그리고 의사결정나무 분석 등 크게 3가지 기법을 적용해 보았다. 그래서 이 3 가지 기법 각각에 같은 조건으로 동일한 데이터를 적용해 결과를 도출해 본 다음, 그 결과들을 비교하여 가장 성과가 우수한 기법을 최종적으로 선택하고자 하였다.

예측의 대상이 되는 변수는 고객이 특정 상품군을 구매한 이력이 있는지, 없는지를 토대로 생성하였으며, 이 때, 상품군의 구분은 G 디아이트 쇼핑몰에서 대분류로 사용하고 있는 4가지 상품군 분류 체계를 그대로 도입해 사용하였다. 각 상품군의 명칭과 특성은 다음의 <표 3>과 같다.

<표 2> 연관규칙 마이닝 적용 결과

신뢰도(%)	지지도(%)	생성된 규칙	해석
48.21	3.75	MDEPART015 → MDEPART001	매직 리프팅 → 매직 마스크
28.95	1.53	MDEHRWE011 → MDEPART007	웨이브 홀라후프 → 파워 세라믹 종아리 디아이트
33.33	1.25	MDGVDCD009 → MDEPART007	김원희의 얼굴이 작아졌어요 VD → 파워 세라믹 종아리 디아이트
47.37	1.25	MDGLIKE008 → MDEPART001	미용 척추 벨트 → 매직 마스크
29.63	1.11	MDGVDCD009 → MDEPART001	김원희의 얼굴이 작아졌어요 VD → 매직 마스크
20.00	0.97	PDFMEAL016 → MDGFASH004	깜짝변신 디아이트 세트 → 디아이트 반바지
33.33	0.97	MDEHRWE007 → MDEHRWE002	만보계 TA5623 → 디지털 체중계 보급형
35.00	0.97	MDEHRWE002 → MDEHRWE007	디지털 체중계 보급형 → 만보계 TA5623
62.50	0.69	MDGVDCD022 → MDGVDCD020	조성아의 메이크업 VD → 이영자의 점프 디아이트 VD
29.41	0.69	MDGVDCD020 → MDGVDCD022	이영자의 점프 디아이트 VD → 조성아의 메이크업 VD
29.41	0.69	MDGVDCD004 → MDGFASH004	다이어트 음악 CD → 다이어트 반바지
38.46	0.69	MDGLIKE001 → MDEPART001	오레그 다리교정 벨트 고급형 → 매직 마스크
35.71	0.69	MDGFASH021 → MDEPART001	헬스 펀닝복 셋트 → 매직 마스크
83.33	0.69	MDEHRWE004 → MDEPART007	기계식 체중계 고급형 → 파워 세라믹 종아리 디아이트

로지스틱회귀분석과 의사결정나무의 경우, 모형구축을 위한 데이터는 전체 데이터의 80%를 사용하였으며, 검증용 데이터는 나머지 20%를 사용하였다. 한편 과적합(overfitting) 현상을 막기 위해 시험용 데이터를 별도로 필요로 하는 인공신경망의 경우에는 모형구축 : 시험 : 검증용 데이터가 각각 60% : 20% : 20%의 비중이 되도록 설계하여 실험하였다.

로지스틱회귀분석의 경우, 단계별 로지스틱회귀분석(stepwise logistic regression) 모형을 활용하였는데, 단계선택방법은 전진-조건(forward-conditional), 단계선택기준은 진입시 0.05, 제거시 0.10으로, 분류기준값은 결과값의 0.5를 기준으로 하였다. 인공신경망 모형은 3층 구조의 은닉층이 1개인 네트워크 모형을 기준 모형으로 활용하였는데, 은닉층 노드의 개수는 입력변수와 출력변수 개수의 합을  $n$ 이라 할 때,  $n/2$ ,  $n$ ,  $3n/2$ ,  $2n$ 의 총 4가지 경우를 대상으로 실험하였다. 또한 학습율과 관성율은 0.1로, 학습중단기준은 시

〈표 3〉 상품군 분류 체계

번호	상품군에 대한 설명
1	다이어트 클리닉 상품군 : 제품과 상담 서비스가 하나의 프로그램 형태로 제공되는 다이어트 종합 솔루션 제공 상품으로서, 고가 제품이 주류임
2	다이어트 식품 상품군 : 식사 대신 먹을 수 있는 다이어트 식품이나 다이어트에 도움을 주는 기타 보조 식품들이 여기에 해당됨
3	다이어트 운동 상품군 : 다이어트를 위한 각종 운동 기구들, 축정 도구들이 해당되는 상품군
4	다이어트 용품 상품군 : 기타 다이어트에 활용될 수 있는 액세서리, 도서 및 음반, 패션 용품 등 나머지 부류의 다이어트 보조 기구들이 이 상품군에 해당됨

험용 자료에서 최소 오류율을 기록한 이후로 20,000번의 학습을 허용하였다. 의사결정나무 분석은 가장 널리 활용되고 있는 C4.5 알고리즘을 적용하였다. 이 때, 각 노드로부터 뺀어나올 수 있는 가지의 최대값은 2로 설정하였고, 나무의 최대깊이는 6으로 설정한 뒤, 실험 결과를 확인하였다. 이상의 세 모델을 적용한 결과, 최종적으로 얻게 된 실험 결과는 다음의 〈표 4〉와 같다.

하단의 결과에서 볼 수 있듯이, 본 연구에서 적용해 본 사례에서는 각 상품군별 구매행동 예측모형으로서 의사결정나무의 예측력이 가장 높게 나오는 것으로 나타났다. 그리하여, G 온라인 쇼핑몰을 위한 상품 추천 시스템의 모듈 2는 의

사 결정 나무를 통해 추천하는 형태로 구성하게 되었다. 이처럼 본 사례에서 다른 기법들에 비해 의사결정나무의 예측력이 높은 결과를 보인 것은 사용된 데이터셋이 상대적으로 ‘비율형 변수’보다는 주로 ‘범주형 변수’를 많이 포함하고 있는 점에 기인하고 있는 것으로 추측된다

#### 4.2.3 조정 에이전트를 이용한 상품 추천

조정 에이전트에서는 앞서 구축한 연관관계 분석의 결과를 토대로 얻어지는 모듈 1의 추천 결과와 의사결정나무에 의한 모듈 2의 추천 결과를 모아 우선 순위를 조정하고, 최종적으로 고객에게 추천 결과를 보여주는 역할을 수행한다. 만약 모듈 1과 모듈 2에서 아무런 추천 결과를 내지 않을 경우에는 조정 에이전트가 자동으로 전체 구매 데이터를 탐색해 전체적으로 가장 많이 팔린 상품을 중심으로 추천 결과를 생성하게 된다.

아울러 조정 에이전트는 모듈 2의 추천 결과, 즉 상품군 단위의 추천 결과를 구체적인 상품 단위의 추천 결과로 전이하는 일을 수행하게 된다. 이는 판매자가 어떤 상품을 보이게 할 것인지를 미리 선택해 놓은 결과를 토대로 처리하게 된다. 또한 조정 에이전트는 모듈 2에서 여러 개의 상품군이 선택된 경우, 상대적으로 단가가 높고 판매자에게 유리한 상품군을 먼저 소개하는 기능도 수행한다. 본 사례의 경우에는 실제 G 쇼핑몰의 특성을 반영하여, ‘상품군 1 (다이어트 클리닉) > 상품군 2 (다이어트 식품) > 상품군 3 (다이어트 운동) > 상품군 4 (다이어트 용품)’의 순서대로 상품 추천의 우선 순위를 결정하도록 하였다.

〈표 4〉 전체 분류모형들의 실험 결과

상품군	모형구축 데이터			검증용 데이터		
	로지스틱분석	인공신경망	의사결정나무	로지스틱분석	인공신경망	의사결정나무
1	75.32%	77.90%	84.47%	67.80%	72.88%	80.51%
2	64.99%	65.92%	67.14%	61.43%	60.95%	62.38%
3	58.94%	58.21%	62.39%	59.65%	60.44%	65.66%
4	57.55%	57.37%	63.26%	56.25%	56.45%	57.46%

### 4.3 프로토타입 구축

앞서 다양한 실험을 통해 설계된 추천 모델이 과연 실제적으로 적용했을 때, 어떤 결과를 보일 것이며, 또 모델을 통해 도출된 추천 결과에 이용자들은 어떤 평가를 내릴 것인지 확인해보기 위해, 본 연구에서는 구축된 모델을 프로토타입

형태의 웹 기반 시스템으로 구축하였다. 완성된 프로토타입은 아래 <그림 3>과 같은 화면에서 방문객이 자신의 정보를 입력하고, <그림 4>의 화면에서 자신이 예전에 구매한 경험이 있는 상품을 입력하면, 최종적으로 <그림 5>와 같은 화면을 통해 최종적인 추천 결과를 제시하는 형태로 설계되어 있다.

The screenshot shows a Microsoft Internet Explorer window with the title bar "포문 Gooddiet.com : Product Recommender System : Prototype Ver 1.0 - Microsoft Internet Explorer". The main content area displays a form titled "Product Recommender System using DataMining". The form asks for user information such as gender (Male or Female), birth year (1976), and height (Cm or Kg). It also includes a section for dietary habits (Vegetarian, Vegan, etc.) and a note about dietary supplement preferences. At the bottom, there is a note asking if the user has ever used the service before.

<그림 3> 프로필 입력 화면



<그림 4> 구매 이력 입력 화면



&lt;그림 5&gt; 최종 추천 결과 제시 화면

#### 4.4 온라인 설문을 통한 유용성 검증

본 연구에서는 앞서 제안하고자 한 추천 모델의 유용성을 검증하기 위해 Krulwich(1997)의 연구에서 적용했던 방법처럼, 프로토타입 시스템에 온라인 설문 모듈을 첨부해 추천 결과의 유용성에 대한 사용자들의 평가를 받아보았다. 제안 모델의 유용성을 보다 정확하게 검증하기 위한 본 설문은 크게 2단계로 나누어 이루어졌는데, 1단계에서는 본 연구의 추천 알고리즘을 통해 상품을 추천한 뒤, 그 추천 결과에 대한 사용자들의 평가를 입력 받았고, 2단계에서는 무작위로 상품을 추천해 보고, 그 추천 결과에 대한 사용자들의 평가를 입력 받아 보았다.

이 때 무작위 결과와 추천 알고리즘을 통한 결과 중 어떤 추천 결과를 먼저 보여줄 것인가 하는 것은 실험의 공정성을 기하기 위해, 컴퓨터가 무작위로 선정하도록 하였다. 그리고, 사용자가 입력하는 평가 결과는 총 5점 리커트(Likert) 척도로 입력하도록 하였다. <표 5>는 설문조사의 결과를 정리한 것이다.

<표 5>에서 볼 수 있듯이, 추천시스템을 통해

추천된 결과가 무작위로 추천된 결과에 비해 근소한 차이로 정보의 유용성 측면에서 앞서는 것으로 나타났다. 특히 누적 확률 분포로 미루어 볼 때, 추천 시스템을 통해 추천된 결과가 특정 고객에게 높은 점수를 받은 비중이 무작위로 추천한 결과에 비해 상대적으로 높은 것으로 나타났음을 알 수 있다.

&lt;표 5&gt; 온라인 설문의 결과

	추천시스템 결과		무작위 결과		
	빈도수	누적확률	빈도수	누적확률	
점수	1	0	0.00%	8	10.81%
	2	14	18.92%	13	28.38%
	3	33	63.51%	37	78.38%
	4	19	89.19%	15	98.65%
	5	8	100.00%	1	100.00%
평균점수	3.284		2.824		

두 만족도 점수 차이가 통계적으로 유의한지를 확인하기 위해 통계적 검증을 적용해 보았다. 통계 소프트웨어인 SPSS for Windows 13.0을 이용해 ‘대응표본 T검정(paired-samples T-test)’을

〈표 6〉 온라인 설문결과에 대한 통계적 검증결과

	Paired Differences					t	df	Sig. (2-tailed)			
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference							
				Lower	Upper						
SC1-SC2	.446	1.251	.145	.156	.736	3.066	73	.003			

수행하였는데, 그 결과가 다음의 〈표 6〉에 제시되어 있다. 수행 결과, 근사 유의확률이 0.003로 나타나 95% 신뢰수준 하에서 두 만족도 점수간 평균의 차이는 통계적으로 유의함을 확인할 수 있었다.

## V. 결 언

본 연구에서는 데이터 마이닝 기법과 애이전트의 개념을 도입함으로서, 기존 상품추천방식들의 한계점을 어느 정도 극복할 수 있는 새로운 형태의 상품추천시스템을 제안하였으며, 이를 실제 국내 인터넷 쇼핑몰 사례에 적용하고, 그 시스템의 만족도를 설문 조사해 봄으로서, 시스템의 유용성을 검증하였다. 본 연구가 지니는 의의는 크게 다음의 2가지로 제시될 수 있다.

첫째는 본 연구가 인터넷 쇼핑몰이 기획보한 고객 정보를 최대한 활용하여, 구매 이력이 전혀 없는 고객에게도 효과적인 추천을 할 수 있는 새로운 추천 알고리즘을 제시하고 있다는 점을 들 수 있다. 특히 본 연구의 사례적용결과를 통해 미루어 볼 때, 상품군 1의 예측율이 최대 80% 이상까지 될 수 있다는 사실은 특정 분야에 전문성을 갖는 중·소형 인터넷 쇼핑몰의 경우, 이러한 프로필 기반의 추천 방식이 상당히 효과적으로 적용될 수 있음을 입증하고 있다고 할 수 있다.

두 번째는 프로토타입 제시를 통해 모델의 적용 가능성을 실증적으로 입증했다는 점이다. 특히 본 연구에서 직접적으로 측정하지는 않았지만, 개발된 프로토타입의 경우 Pentium III 프로

세서 기반의 PC급 서버 환경 하에서도 거의 실시간으로 고객에게 추천결과를 제시해 줌으로서, 본 연구에서 제시하는 추천시스템이 상당히 적은 연산 용량만을 필요로 함을 실증적으로 입증할 수 있었다.

그 밖에 본 연구의 추천시스템이 상품군의 개념을 도입함으로서 판매자의 의도를 어느 정도 반영할 수 있도록 설계했다든가, 다양한 알고리즘을 실제 온라인 다이어트 쇼핑몰에 적용해 봄으로서 부가적으로 국내 온라인 다이어트 쇼핑몰 이용 고객의 구매행동 패턴을 찾아볼 수 있었다는 점 역시 중요한 의의라고 할 수 있다.

하지만 본 연구의 한계점도 다수 존재한다. 우선, 모델을 구축하는데 있어서, 프로필 정보는 활용하고 있으나, 웹 로그 등의 기타 행태적 정보를 배제시켜 제한적인 정보만을 활용하고 있는 점은 본 연구의 첫번째 한계점이라 하겠다. 또한, 사례 적용에 있어 적은 표본수와 너무 광범위한 개념의 상품군 설정으로 인해 상품군 1, 2를 제외한 나머지 상품군에서 분류 모형의 예측력이 전반적으로 낮게 나타난 점도 본 연구에서 미흡하게 진행된 부분이라고 판단된다. 그리고 제시되고 있는 모델에서 조정 애이전트의 역할이 다소 수동적이고, 단순하게 설계되었다는 점은 본 연구의 또 다른 한계점이라 할 수 있다. 이러한 한계점들을 극복하고, 본 연구에서 달성하고자 했던 목표처럼 수많은 중소규모 인터넷 쇼핑몰들이 어려움 없이 도입할 수 있도록 하는 추천시스템에 대한 연구가 추후 지속적으로 이루어져야 할 것이다.

## 참 고 문 헌

- 강부식, “자기 조직화 신경망을 이용한 협력적 여과 기법의 웹 개인화 시스템에 대한 연구”, *한국지능정보시스템학회논문지*, 제9권, 제3호, 2003, pp. 117-135.
- 김재경, 서지혜, 안도현, 조윤호, “A Personalized Recommendation Methodology based on Collaborative Filtering”, *한국지능정보시스템학회 논문지*, 제8권, 제2호, 2002, pp. 139-157.
- 김재경, 안도현, 조윤호, “Development of a Personalized Recommendation Procedure based on Data Mining Techniques for Internet Shopping malls”, *한국지능정보시스템학회논문지*, 제9권, 제3호, 2003, pp. 177-191.
- 김재경, 안도현, 조윤호, “개인별 상품추천시스템, WebCF PT: 웹마이닝과 상품계층도를 이용한 협업필터링”, *경영정보학연구*, 제15권, 제1호, 2005, pp. 63-79.
- 김종우, 배세진, 이홍주, “협업 필터링 기반 개인화 추천에서의 평가자료의 회소 정도의 영향”, *경영정보학연구*, 제14권, 제2호, 2004, pp. 131-149.
- 조윤호, 박수경, 안도현, 김재경, “재구성된 제품 계층도를 이용한 협업 추천 방법론 및 그 평가”, *한국경영과학회지*, 제29권, 제2호, 2004, pp. 59-75.
- Adomavicius, G. and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, 2005, pp. 734-749.
- Aggarwal, C.C., Z. Sun, and P.S. Yu, “Online Algorithms for Finding Profile Association Rules”, *Proceedings of the seventh international conference on Information and knowledge management*, 1998, pp. 86-95.

- Balabanovic, M. and Y. Shoham, “Fab: Content Based, Collaborative Recommendation”, *Communications of the ACM*, Vol.40, No.3, 1997, pp. 66-72.
- Berry, M.J.A. and G. Linoff, *Data Mining Techniques: For Marketing, Sales and Customer Support*, New York, NY, John Wiley & Sons, 1997.
- Cho, Y.H., J.K. Kim, and S.H. Kim, “A Personalized Recommender System based on Web usage Mining and Decision Tree Induction”, *Expert Systems with Applications*, Vol.23, 2002, pp. 329-342.
- Cho, Y.H. and J.K. Kim, “Application of Web usage Mining and Product Taxonomy to Collaborative recommendations in e-commerce”, *Expert Systems with Applications*, Vol.26, 2004, pp. 233-246.
- Funakoshi, K. and T. Ohguro, “A Content Based Collaborative Recommender System with Detailed Use of Evaluations”, *Proceedings of the 4th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies*, 2000, pp. 253-256.
- Goldberg, D., D. Nichols, B.M. Oki, and D. Terry, “Using Collaborative Filtering to Weave an Information Tapestry”, *Communications of the ACM*, Vol.35, No.12, 1992, pp. 61-70.
- Iyengar, V. and T. Zhang, “Empirical Study of Recommender Systems Using Linear Classifiers”, *Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Apr. 2001, pp. 16-27.
- Kim, C.Y., J.K. Lee, Y.H. Cho, and D.H. Kim, “VISCORS: A Visual Content Recommender for the Mobile Web”, *IEEE Intelligent Systems*, Vol.19, Issue 6, 2004, pp. 32-39.
- Kim, D. and B.J. Yum, “Collaborative Filtering

- based on Iterative Principal Component Analysis”, *Expert Systems with Applications*, Vol. 28, No.4, 2005, pp. 823-830.
- Kim, J.K., Y.H. Cho, W.J. Kim, J.R. Kim, and J.H. Suh, “A Personalized Recommendation Procedure for Internet Shopping Support”, *Electronic Commerce Research and Applications*, Vol.1, 2002, pp. 301-313.
- Kim, K.-S. and I. Han, “The Cluster-Indexing Method for Case based Reasoning using Self-organizing Maps and Learning Vector Quantization for Bond Rating Cases”, *Expert Systems with Applications*, Vol.21, No.3, 2001, pp. 147-156.
- Konstan, J., B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, “GroupLens: Applying Collaborative Filtering to Usenet News”, *Communications of the ACM*, Vol.40, No.3, Mar. 1997, pp. 77-87.
- Krulwich, B., “Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data”, *Artificial Intelligence Magazine*, Vol. 18, Issue 2, Summer 1997, pp. 37-45.
- Lawrence, R., G. Almasi, V. Kotlyar, M. Viveros, and S. Duri, “Personalization of Supermarket Product Recommendations”, *Data Mining and Knowledge Discovery*, Vol.5, Issue 1-2, Jan.-Apr. 2001, pp. 11-32.
- Lee, C.H., Y.H. Kim, and P.K. Rhee, “Web Personalization Expert with Combining Collaborative Filtering and Association Rule Mining Technique”, *Expert Systems with Applications*, Vol. 21, Issue 3, 2001, pp. 131-137.
- Liu, D.-R. and Y.-Y. Shih, “Integrating AHP and Data Mining for Product Recommendation based on Customer Lifetime Value”, *Information & Management*, Vol.42, No.3, 2005, pp. 387-400.
- Pazzani, M.J., “A Framework for Collaborative, Content-based and Demographic Filtering”, *Artificial Intelligence Review*, Vol.13, No. 5-6, 1999, pp. 393-408.
- Resnick, P., N. Iacovou, M. Suchak, and P. Bergstrom, “GroupLens: An Open Architecture for Collaborative Filtering of Netnews”, *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1994, pp. 175-186.
- Roh, T.H., K.J. Oh, and I. Han, “The Collaborative Filtering Recommendation based on SOM Cluster-Indexing CBR”, *Expert Systems with Applications*, Vol.25, No.3, 2003, pp. 413-423.
- Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, “Application of Dimensionality Reduction in Recommender Systems: A Case Study”, *Proceedings of the WebKDD Workshop at the ACM SIGKDD*, 2000.
- Schafer, J., J. Konstan, and J. Riedl, “E-Commerce Recommendation Applications”, *Journal of Data Mining and Knowledge Discovery*, Vol.5, No.1-2, 2001, pp. 115-153.
- Stegmann, R., M. Koch, M. Lacher, T. Leckner, and V. Renneberg, “Generating Personalized Recommendations in a Model-Based Product Configurator System”, *Proceedings of Workshop on Configuration held in conjunction with 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, Mexico, Aug. 9-15, 2003.
- Wang, Y.-F., Y.-L. Chuang, M.-H. Hsu and H.-C. Keh, “A Personalized Recommender System for the Cosmetic Business”, *Expert Systems with Applications*, Vol.26, No.3, 2004, pp. 427-434.
- Wielenga, D., B. Lucas and J. Georges, *Enterprise Miner(TM): Applying Data Mining Techniques Course Note*, SAS Institute Inc., Cary, NC,

1999.

Wu, K.-L., C.C. Aggarwal, and P.S. Yu, "Personalization with dynamic profiler", *Proceedings of the Third International Workshop on Advanced Issues of E-commerce and Web-based Infor-*

*mation Systems*, 2001, pp. 12-20.

Yu, P.S., "Data Mining and Personalization", *Proceedings of the Sixth International Conference on Database Systems for Advanced Applications*, April 19-21, 1999, pp. 6-13.

Information Systems Review

Volume 8 Number 1

April 2006

## The Product Recommender System Combining Association Rules and Classification Models: The Case of G Internet Shopping Mall

Hyunchul Ahn\* · Ingoo Han\* · Kyoung-Jae Kim\*\*

### Abstract

As the Internet spreads, many people have interests in e-CRM and product recommender systems, one of e-CRM applications. Among various approaches for recommendation, collaborative filtering and content-based approaches have been investigated and applied widely. Despite their popularity, traditional recommendation approaches have some limitations. They require at least one purchase transaction per user. In addition, they don't utilize much information such as demographic and specific personal profile information. This study suggests new hybrid recommendation model using two data mining techniques, association rule and classification, as well as intelligent agent to overcome these limitations. To validate the usefulness of the model, it was applied to the real case and the prototype web site was developed. We assessed the usefulness of the suggested recommendation model through online survey. The result of the survey showed that the information of the recommendation was generally useful to the survey participants.

**Keywords:** Customer Relationship Management, Recommender Systems, Data Mining, Association Rule, Intelligent Agent

---

\* Graduate School of Management, Korea Advanced Institute of Science and Technology

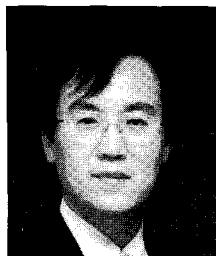
\*\* Department of Management Information Systems, Dongguk University

## ● 저 자 소 개 ●



안 현 철 (hcahn@kaist.ac.kr)

현재 한국과학기술원 테크노경영대학원 박사과정에 재학 중이다. KAIST에서 산업경영학사 및 경영공학석사를 취득하였다. 주요 관심분야는 인공지능 및 데이터마이닝을 이용한 재무예측, 고객관계관리, m-CRM 등이다.



한 인 구 (ighan@kgsm.kaist.ac.kr)

현재 한국과학기술원 테크노경영대학원 교수로 재직 중이다. 서울대학교 국제 경제학사, KAIST 경영과학석사를 취득하였고, University of Illinois at Urbana-Champaign에서 회계정보시스템을 전공하여 경영학박사를 취득하였다. 주요 관심분야는 지능형 신용평가시스템, 인공지능을 이용한 재무예측, 지식자산 가치평가, 정보시스템 감사 및 보안 등이다.



김 경 재 (kjkim@dongguk.edu)

현재 동국대학교 경영대학 경영정보학과 교수로 재직 중이다. 중앙대에서 경영학사를, KAIST에서 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 주요 관심분야는 데이터마이닝, 지능형 에이전트, 고객관계관리, 지식경영 등이다.

논문접수일 : 2005년 8월 31일  
1차 수정일 : 2006년 2월 10일

게재확정일 : 2006년 2월 23일