

# 사례기반추론과 규칙기반추론을 이용한 e-쇼핑몰의 상품추천 시스템

이 건 호<sup>†</sup> · 이 동 훈<sup>††</sup>

## 요 약

e-쇼핑몰 경영자들은 고객들의 다양한 제품 구매 욕구를 충족시키기 위한 효율적 시스템에 많은 관심을 가지고 있다. 인터넷 쇼핑몰 운영에 있어 고객들의 개인적 구매 특성 및 취향을 파악하여 고객들을 효과적으로 관리 하는데 많은 어려움이 있다. 상품 추천의 과정이 기획된 소수의 특정 상품을 고객의 유형 및 특성들의 고려 없이 공급자 중심으로 이루어져 고객관리의 문제점으로 지적되고 있다. 본 연구에서는 고객위주의 추천을 위해 규칙기반추론(Rule-Based Reasoning, RBR)과 사례기반추론(Case-Based Reasoning, CBR)을 하여 고객의 취향 및 구매 특성에 따른 추천방법을 제시한다. 기존의 제품 판매정보와 고객정보를 이용해 생성한 규칙베이스와 사례베이스의 고객특성과 입력된 고객특성의 유사도를 평가해서 고객의 취향에 따라 추천하도록 한다. 생성된 규칙과 사례기반의 추론으로 기존의 정보를 효과적으로 사용하고 또한 고객 및 시장 상황의 변화를 인식하고 지속적인 학습을 수행하여 지능적 추천이 이루어진다.

## Recommending System of Products on e-shopping malls based on CBR and RBR

Gun-Ho Lee<sup>†</sup> · Dong-Hun Lee<sup>††</sup>

## ABSTRACT

It is a major concern of e-shopping mall managers to satisfy a variety of customer's desire by recommending a proper product to the perspective purchaser. Customer information like customer's fondness, age, gender, etc. in shopping has not been used effectively for the customers or the suppliers. Conventionally, e-shopping mall managers have recommended specific items of products to their customers without considering thoroughly in a customer point of view. This study introduces the ways of a choosing and recommending of products using case-based reasoning and rule-based reasoning for customer themselves or others. A similarity measure between one member's idiosyncrasy and the other members' is developed based on the rule base and the case base. The case base is improved for the system intelligence by recognizing and learning the changes of customer's desire and shopping trend.

**키워드 :** 추천시스템(Recommending System), 사례기반추론(Case-based Reasoning), 규칙기반추론(Rule-based Reasoning) 개인화(Personalization)

### 1. 서 론

인터넷을 통한 전자상거래가 급격히 늘어남에 따라 제품에 대한 정보량이 증가하여 고객의 정보 욕구는 충족시키지 못, 고객에게 불필요하거나 너무 많은 정보가 제공되어 고객 스스로 이러한 정보를 이용하기보다는 정보의 공해에 시달리고 있기도 하다. 따라서 고객에게 적합한 정보를 분석하여 선별적으로 제공할 필요가 있다[1].

인터넷이 보편화됨에 따라 인터넷 쇼핑몰에서의 모든 거래 행위는 데이터베이스로 축적할 수 있게 되었고 고객의 모든 정보를 이용, 분석하여 해당 고객에게 적합한 서비스를

선별적으로 제공하는 것이 가능해진 것이다[2].

최근 전자 상거래에서 고객들의 만족도 증가, 효율적인 고객관리, 마케팅 비용의 절감, 서비스 및 제품의 개선 등은 경쟁력 강화를 위해 개인화(Personalization)가 많은 관심의 대상이 되고 있다[8]. 개인화는 고객이 자신의 선호, 관심, 구매경력과 같은 정보를 웹 사이트에 제시하면 웹 사이트는 고객이 제시한 정보를 기초로 고객에게 가장 알맞은 구매정보를 제공하고자 한다. 따라서 e-쇼핑몰 운영자는 회원고객에 관한 정보를 얻고 고객의 지속적인 이용이나 구매를 하도록 유도하며 고객은 자신에게 가장 알맞은 구매정보를 보다 편리한 방법으로 얻을 수 있어야 할 것이다.

이를 위해 연관규칙(association rule)을 이용하여 구매정보를 제공하고자 하는 연구[3, 13]가 있었으나 연관규칙을 이용한 경우 새로운 사례와 일치하는 규칙을 발견하지 못했을

<sup>†</sup> 종신회원 : 숭실대학교 산업·정보시스템공학과 교수  
<sup>††</sup> 준회원 : 숭실대학교 대학원 산업·정보시스템공학과  
논문접수 : 2004년 3월 11일, 심사완료 : 2004년 6월 7일

경우는 어떠한 추천도 할 수 없다는 단점을 내포하고 있다.

연관규칙은 항목 집합으로 표현된 트랜잭션에서 각 항목 간의 연관성을 반영하는 규칙이다. Agrawal, Imielinski와 Swami에 의해 1993년 처음 소개된 후[7], 데이터베이스 검색 횟수를 줄이거나, 주 기억 장치의 한계를 없애는 등의 발전된 알고리즘들이 발표되어 왔다.

또한 사례기반추론(case-based reasoning, CBR)을 활용한 상품추천에 대한 연구[9, 12]에서도 고객의 구매의도와는 상관없이 정확히 고객들의 극히 적은 특성들에만 의존한 추론으로 적절한 제품을 추천하지 못하고 있다. 뿐만 아니라 고객특성의 유사도 측정에서 단순히 일치하면 유사도 값을 부여하고 그렇지 않은 경우에는 부여하지 않는 방식을 사용함으로써 다양한 측면을 고려하지 못하고 있다.

본 연구에서는 고객의 인터넷 쇼핑물 구매의도를 정확히 파악하여 고객의 의도에 맞는 추천을 위해 과거의 구매이력에 의한 연관관계 분석을 통한 규칙기반추론(rule-based reasoning, RBR)과 과거의 경험에 의한 보다 다양한 사례를 이용하여 새로운 문제를 해결하는 CBR기법의 개선된 추천기법을 제안하고자 한다.

CBR은 과거의 어떤 문제를 해결하기 위해 사용했던 경험을 바탕으로 새로운 문제를 해결하는 방법이다. 새로운 문제에 직면했을 때, 이를 해결하는 방법으로 과거의 비슷한 상황을 기억하고 그 상황에서 사용했던 정보와 지식을 재사용함으로써 새로운 문제를 해결하는 것으로 사례조회(REtrieval), 사례재이용(REuse), 사례수정(REvise), 그리고 사례저장(REtain)의 4단계(4REs)를 통해 추론을 하는 것이다[11].

## 2. 관련 연구

상품추천시스템을 위한 기존의 연구들은 연관규칙이나 CBR 혹은 연관규칙과 CBR의 통합을 주요 기법으로 이용하고 있다. 연관규칙에 의한 상품추천시스템에 대한 기존의 연구[13]는 비논리적이거나 학습성을 필요로 하는 경우 연관규칙만으로 상품 추천시스템을 구성하는 데는 추론의 경직성을 갖게 되는 한계를 가지고 있다[15].

CBR은 인공지능기법으로 많은 기존 연구에서 학습성과 사용자 환경에 대한 적응성이 입증되었다[5, 14-16]. CBR을 이용한 개인화 추천시스템에 대한 연구가 있었으나[14] 기존의 사례를 성공적으로 평가하여 유사한 사례를 제시하여야 하나, 이는 기존의 사례가 충분 할 때 만 가능한 것이다. 만약 사례의 연관성이 검색되지 않을 경우 어떠한 추천도 불가능하며 단순한 구매패턴에 의존한 방식이므로 고객의 특성을 고려하지 못한다는 단점이 있다. 사례가 충분하지 않거나 연관규칙의 적용이 가능한 상황이면 연관규칙에 의한 상품 추천을 하게 된다. 위의 두 기법의 한계를 극복하기 위해 연관규칙과 CBR 기법을 이용한 구매지원 시스템을 제시하

기도 하였다[12]. 그러나 CBR에서 가장 중요한 유사도 평가를 허철회 등의 기존연구[10, 12, 17]에서 사용한 방법으로 측정하였으나 사례를 결정하는 있어 다음과 같은 내용은 고려하지 않고 있다.

### 첫째, 상품의 유행성

광범위한 시장에서 다양하고 민첩한 제품의 변화에 따른 사례베이스의 적절성을 고려하여야 한다. 도서와 같은 상품은 시간이 지나도 찾을 수가 있지만, 유행에 민감한 의류와 같은 패션 제품은 일정시간이 지남에 따라 고객의 취향이 달라지는 점을 고려하여야 할 것이다.

### 둘째, 고객 특성별 가중치 부여

유사도 측정에서 고려하는 가중치의 경우 기존의 연구에서는 단순한 경험에 의존하고 있으나 고객의 각 특성별 대표 제품의 판매수와, 각 고객 특성에서 카테고리가 차지하는 정도를 반영하여 고객특성들 중 구매에 미치는 구매경향에 따라 정확히 가중치를 결정해야 한다.

### 셋째, 사례베이스의 다양성

기존의 추천기법에서 사용되는 사례베이스는 고객의 구매의도를 파악하지 않은 상태에서 단순히 고객의 특성과 사례의 유사성을 평가 제품을 추천하고 있다. 그러나 선물의 경우와 같이 구매자와 실제 제품 사용자는 다를 수 있으며 구매자와 제품 사용자와의 관계에 따른 분석과 사회적 관습에 의한 구매사례, 등의 분석 없이 추천한다면 구매자의 의도와 맞지 않게 부적절한 추천이 되는 것이다.

### 넷째, 관심분야의 유사성

e-쇼핑몰의 상품추천에서 가장 중요한 고려되어야 하는 항목의 하나인 관심분야의 분류에 따라 유사성의 강약 정도를 반영되어야 하지 못하고 있다.

따라서 본 연구에서는 위에서 언급한 4가지 문제들을 해결하는 시스템을 개발하고자 한다.

## 3. 추천시스템의 설계

고객특성정보와 판매되는 제품간의 관계 속에서 특성별 제품구매의 연관성을 파악하여 가중치를 부여하는 방법과 제품의 유행에 따른 민감도를 추론에 반영할 수 있도록 최신도를 제시한다. 각 고객 특성에 따른 가중치와 최신도를 고려한 새로운 유사도 평가방법을 제안하며 구매자와 제품실 사용자와의 관계에 따라 적절한 사례를 가질 수 있도록 사례베이스를 설계하고자 한다.

### 3.1 시스템의 구성

제안된 시스템은 사용단계에 따라 다음과 같은 하부시스템으로 구성되어 있다. 첫째 고객, 제품 및 판매정보의 데이

데이터베이스 구조는 (그림 1)과 같다.

제품정보						
제품코드	제품명	제조업체	가격	판매량	제품 등록일	사진URL
고객정보						
ID	성명	주소	지역	Tel	E-mail	생년월일
						성별
						나이
						직업
						결혼여부
						학력
						관심분야
						비밀번호
						가장최근 구입제품
판매정보						
유형	제품코드	고객ID	수량	주문일자	배달일자	지불방법

(그림 1) 제품/고객/판매정보의 데이터베이스구조

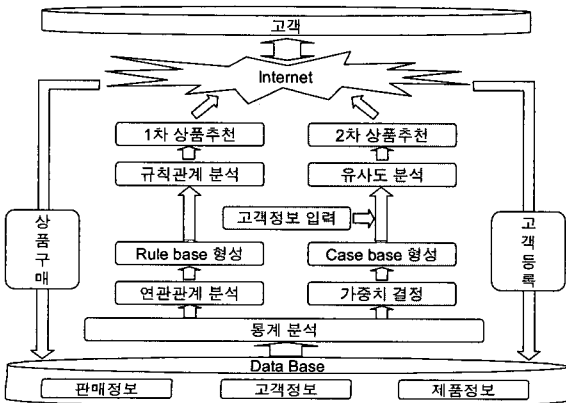
둘째로 데이터베이스의 자료를 정량화해서 통계적 처리가 가능하도록 전처리 과정을 가지며, 셋째로 통계적 분석을 통해 규칙베이스 형성을 위한 연관관계 및 가중치를 계산한다. 이 단계의 규칙베이스는 연관분석을 통한 결과를 사용하는데 이는 Apriori 알고리즘[6]에 입각해 Clementine(ver.6.5)[4]을 이용한다. 다음 단계로 규칙베이스 및 사례베이스를 형성하며, 구조는 (그림 2)와 같다.

사례베이스-1											
사례번호	유형	나이	성별	관심분야	학력	결혼여부	직업	지역	구입제품	성공회수	
사례베이스-2											
사례번호	유형	관계	나이	성별	관심분야	학력	결혼여부	직업	지역	구입제품	성공회수
규칙베이스											
제품A	제품B	총구매수	점유율	규칙정확도							

(그림 2) 지식베이스의 자료구조

사례베이스는 (그림 2)와 같이 자신을 위한 구매를 목적으로 한 고객과 타인을 위한 구매를 목적인 두 가지 유형으로 구분한다.

다섯 번째는 추천대상자의 정보입력 단계이다. 상품 구매자 자신을 위한 경우 새로운 고객인지, 기존의 고객인지에 따라 추론을 한다. 선물을 할 경우 대상자와의 관계 및 특성을 구매지원 추론에 이용하게 된다.



(그림 3) 상품 추천시스템의 구성

여섯 번째로 입력된 정보를 통해 규칙관계 및 사례베이스와의 유사도를 측정하게 되고 다음 단계에서 추론된 제품 정보를 고객에게 추천을 한다.

마지막 단계로 추천된 정보에 대한 고객의 반응을 평가하고 feedback한다. 고객이 기존의 사례에 의해 추천받아 구매를 하면 기존사례의 성공회수를 갱신하고 추천하지 않은 제품을 새롭게 구매할 경우는 새로운 사례베이스 구축과 함께 새로운 규칙베이스를 형성하게 된다.

### 3.2 각 특성의 가중치 계산방법

각 특성 별 가중치란 각 특성별 구매에 영향을 미치는 상대적인 중요도 나타내는 값이다. <표 1>과 같이 쇼핑몰의 주요 제품들의 판매자료에 근거하여 구매회원수가 어떤 카테고리들에 집중되는 구매경향을 이루고 있는 특성에 상대적으로 높은 점수를 부과하여 전체 유사도 평가에서 반영하고자 한다.

본 연구에서는 자료의 통계적 처리에서 그 분포의 산포도 (degree of scattering)를 나타내는 특성량에 해당하는 평균편차(mean deviation)를 이용하여 식 (1)과 같이 각 특성들의 카테고리별 구매자수 분포의 정도를 평가한다.

$$W_i = \frac{\sum_{j=1}^{n_i} |x_{ij} - \bar{x}_i|}{n_i} \times C \quad (1)$$

$i = \{ loc \text{ (지역)}, gen \text{ (성별)}, age \text{ (나이)}, job \text{ (직업)}, mar \text{ (결혼여부)}, edu \text{ (교육정도)}, int \text{ (관심분야)} \dots \}$

$x_{ij}$ : 고객특성  $i$ 의 카테고리  $j$ 의 판매수

$$\bar{x}_i = \sum_{j=1}^{n_i} x_{ij} / n_i$$

$n_i$ : 고객특성  $i$ 의 카테고리의 수

$C$ : 가중치 조정값 상수

<표 1> G 전문 쇼핑몰 구매자의 특성

연령	10~19	20~29	30~39	40~49	50~59	60~69	70이상
구매회수	0	2	10	24	32	14	3
성별	남			여			
구매회수	45			40			
교육정도	중졸이하	고졸	대졸	대학원이상			
구매회수	5	16	39	25			
소득수준	1500 이하	1501~2000	2001~3000	3001~4000	4001 이상		
구매회수	0	9	17	27	32		
결혼여부	미혼		기혼		이혼		
구매회수	21		42		22		
직업	학생	사무직	기술직	서비스 판매직	생산직	의료인	교육
구매회수	2	16	10	10	3	32	12
관심분야	컴퓨터	정치, 시사	패션, 연예	레저, 여행	부동산	건강, 요리	교육, 학술
구매회수	6	39	6	4	18	1	11

몇몇 카테고리에 구매자수가 집중되어 있으면 산포도가 작다는 의미이며 따라서 평균편차 값도 증가하고 해당 특성에 더 높은 가중치를 갖게 되는 것이다. 가중치는 식 (1)과 같이 가중치의 조정값 상수 C를 사용자가 임의로 결정하도록 한다. <표 1>은 G전문 쇼핑몰의 구매자들의 분포를 나타내고 있다.

age	gender	interest	edu	marital	job	location	product	
1	23	Male	컴퓨터/인터넷	University (Gradu	Single	대학생	서클	1
2	28	Male	스포츠	University (Gradu	Single	사무직	인천	1
3	37	Female	쇼핑	High_school Gra	Married	기술직	강원도	0
4	18	Male	컴퓨터/인터넷	High_school Gra	Single	고등학생	대전	1
5	27	Female	컴퓨터/인터넷	College (Graduat	Married	사무직	경기도	0
6	25	Male	과학/기술	College (Graduat	Single	무직	인천	0
7	42	Male	시사/뉴스	College (Graduat	divorced	기술직	부산	1
8	24	Male	레저/여행	University (Gradu	Single	무직	서울	0
9	21	Male	컴퓨터/인터넷	University (Gradu	Single	대학생	부산	1
10	52	Female	시사/뉴스	University (Gradu	Married	사비스/판매직	광주	1
11	17	Male	컴퓨터/인터넷	High_school Gra	Single	고등학생	대전	0
12	26	Female	과학/기술	University (Gradu	Married	대학원생	충청도	0
13	25	Female	과학/기술	College (Graduat	Single	대학원생	대구	0
14	19	Female	스포츠	High_school Gra	Single	고등학생	울산	0
15	38	Male	정치/사회/종교	College (Graduat	Married	정보통신 관련직	인천	1
16	31	Male	시사/뉴스	College (Graduat	Married	사무직	대전	1
17	22	Female	쇼핑	University (Gradu	Single	대학생	강원도	1
18	20	Female	스포츠	University (Gradu	Single	대학생	광주	0
...	...	...	...	...	...	...	...	...

(그림 4) 고객특성과 대표제품 데이터베이스

식 (1)을 이용한 가중치를 계산하기 위해 C=0.1로 두면,  $W_{연령} = 0.96$ ,  $W_{성별} = 0.25$ ,  $W_{학력} = 1.08$ ,  $W_{소득} = 1.00$ ,  $W_{결혼} = 0.91$ ,  $W_{직업} = 0.68$ ,  $W_{관심분야} = 0.93$ 이 된다. 이 경우 대학 졸 및 대학원졸업자 위주로 학력별로 구매자가 가장 뚜렷하게 나타나고 있고, 다음으로 소득수준 2001 이상의 해당 구매자가 구매경향을 보이고 있는 반면, 성별에 의해서는 거의 차이가 없음을 나타내고 있다.

3.3 사례의 최신도(Newest rate : NR)

소비자의 다양한 요구와 유행에 민감한 반응을 고려하기 위해 식 (2)를 제안한다. 최근의 제품을 대상으로 생성된 사례일수록 최신도 값을 높게 주고 오래된 제품을 구입한 사례의 경우 상대적으로 최신도 값을 낮게 하여 민감한 유행에 대응할 수 있도록 한다.

$$NR_j = 1 - \frac{(Y_c - Y_j) \times 12 + (M_c - M_j)}{S} \quad (2)$$

- $Y_c$  : 추천 시점의 년
- $Y_j$  : 사례 j의 추천제품이 등록된 년
- $M_c$  : 추천 시점의 월
- $M_j$  : 사례 j의 추천제품이 등록된 월
- S : 민감도 지수

식 (2)에서 제시한 민감도 지수는 수명주기(life cycle)에 따라 사용자가 정의할 수 있다. 제품의 변화가 민감할수록 그 수치가 작아지며, 상품이 시간의 흐름에 크게 민감하지 않은 경우는 그 값이 커야 한다. 예를 들어 의류제품의 경우 고객이 2004년 3월에 로그인 했고 민감도 지수가 50이라고 가정하면 검색된 사례가 2003년 12월과 2003년 2월의 사례가 존재한다고 하면, 최신도는 각각 0.94와 0.74을 얻어 오래된 사례가 상대적으로 낮은 유사도 값을 갖게 된다.

3.4 가중치를 부여한 유사도

본 연구에서 고객특성과 최신도를 고려한 고객과 사례의 가중치를 부여한 유사도(weighted similarity measure)를 측정하는 식 (3)는 각 고객특성별 유사도  $Sim(f_i^N, f_i^C)$ 의 가중평균값(weighted average)과 성공정도의 합에 최신도의 곱으로 이루어진다.  $Sim(f_i^N, f_i^C)$ 의 가중평균값은 수학적 기대값(mathematical expectation)을 구하는 정리[18]에 착안하여 각 고객 특성의 가중치와 총 고객특성들의 가중치 합을 비를 각 고객특성에 대한 확률밀도로 간주하여 각 고객 특성의 유사도 값을 각 사건의 값으로 보고 특정사례에 대한 고객특성별 유사도의 가중 평균값을 구하도록 한다.

$$SM = \left( \frac{\sum_{i=age}^{loc} \{W_i \times Sim(f_i^N, f_i^C)\}}{\sum_{i=age}^{loc} W_i} + CS_j \right) \times NR_j \quad (3)$$

- $i = \{age, gen, int, edu, mar, job, loc\}$
- j : 사례번호
- n : 사례수
- $W_i$  : 고객특성 i에 대한 가중치
- $f_i^N$  : 추천대상 고객 특성 i의 특성값
- $f_i^C$  : 사례베이스상의 고객 특성 i의 특성값

$$CS_j = \frac{S_j}{\max\{S_j, j=1, 2, \dots, n\}} \times A$$

해당사례 j의 성공정도( $CS_j$ )에서 성공회수( $S_j$ )는 사례 j가 선정된 횟수이고 성공회수의 최대값은 사례베이스상에 있는 성공회수 중 최대값이며 A는 사례선정 값의 평균값이다[10].

식 (3)의  $Sim(f_i^N, f_i^C)$ 는 각 고객과 사례와의 특성간에 따라 유사도 값을 평가하는 식으로 이들에 대한 측정방법은 아래와 같다.

$$Sim(f_{age}^N, f_{age}^C) = \begin{cases} 0 & \text{IF } |\text{고객나이} - \text{사례베이스나이}| \geq 50 \\ 1 - \frac{|\text{고객나이} - \text{사례베이스나이}|}{50} & \text{OTHERWISE} \end{cases}$$

나이는 실제나이를 이용하여 계산하며 나이가 고객의 나이와 사례의 나이차가 50살 이상이면 유사도가 없는 것으로 한다.

$$Sim(f_{gen}^N, f_{gen}^C) = 1 - |\text{고객성별} - \text{사례베이스성별}|$$

성별은 2가지 카테고리(Male, Female)로 구성되며 같은 성별이면 1의 값을 갖고 다른 성별일 경우는 0의 값을 갖게 한다.

$$Sim(f_{mar}^N, f_{mar}^C) = 1 - |\text{고객결혼여부} - \text{사례결혼여부}|$$

결혼여부는 2가지 카테고리(Single, Married)로 구성된다.

$$\bullet \text{Sim}(f_{int}^N, f_{int}^C) = \begin{cases} 1 & \text{IF 일치하면} \\ 0.5 & \text{ELSE IF } | \text{고객관심분야} - \text{사례관심분야} | \leq 3 \\ 0 & \text{OTHERWISE} \end{cases}$$

관심분야의 카테고리는 <표 2>와 같고, 일치하면 1, 같은 그룹 내에 속할 경우 0.5의 유사도를 부여한다. <표 2>과 같이 회원의 관심분야를 다양하게 설정하더라도 관련성이 있는 항목들을 그룹화 함으로써 정확히 일치하지 않더라도 그룹 내에서는 약간의 유사성을 인정할 수 있도록 유연성을 부여한다.

$$\bullet \text{Sim}(f_{edu}^N, f_{edu}^C) = 1 - \frac{| \text{고객학력} - \text{사례베이스학력} |}{5}$$

학력은 6가지의 카테고리(초졸, 중졸, 고졸, 초대졸, 대졸, 대학원이상)로 구성된다.

$$\bullet \text{Sim}(f_{job}^N, f_{job}^C) = \begin{cases} 1 & \text{IF 일치하면} \\ 0 & \text{OTHERWISE} \end{cases}$$

직업의 경우 19가지 카테고리로 구성되며, 일치하면 1, 그렇지 않으면 0의 유사도를 부여한다.

$$\bullet \text{Sim}(f_{loc}^N, f_{loc}^C) = \begin{cases} 1 & \text{IF 일치하면} \\ 0.5 & \text{ELSE IF } | \text{고객지역} - \text{사례지역} | \leq 3 \\ 0 & \text{OTHERWISE} \end{cases}$$

<표 2> 관심 분야

NO.	그 룹
1	컴퓨터/인터넷
2	과학/기술
6	정치/사회/종교
7	시사/뉴스
11	연예/오락
12	쇼 핑
13	사진/패션
17	벤처창업/경영
18	채테크/증권
22	교육/학술
23	외국어
27	레저/여행
28	스포츠
32	음악/공연
33	영화/비디오
34	문화/예술
35	만화/애니메이션
39	생활/건강
40	음식/요리

<표 3> 지 역

NO.	그 룹
1	서 울
2	인 천
3	경기도
7	부 산
8	대 구
9	울 산
10	경상도
14	광 주
15	전라도
19	대 전
20	충청도
24	강원도
28	제주도

지역의 카테고리는 <표 3>와 같이 구성되며, 유사도는 위의 관심분야와 같이 그룹화를 통해 값을 부여한다.

3.5 고객의도에 따른 유사도 측정

고객의 구매가 선물이 목적인 경우 구매자뿐만 아니라 제품사용자의 특성도 고려할 수 있도록 사례베이스를 설계한다.

자신을 위한 구매일 경우 구매자의 특성정보와 사례베이스-1의 정보를 활용해 유사도를 측정하며, 타인을 위한 구매이면 추가질의를 수행하고 제품을 사용자의 특성을 얻는다. 구매자와 사용대상자의 고객특성은 (그림 2)의 사례베이스-2의 정보와 유사성을 측정한다. 측정방법은 구매자의 고객특성정보와 사례베이스의 구매자의 특성을 식 (3)를 이용해 유사도  $S_G$ 를 계산하고, 제품사용자의 특성정보와 사례베이스의 사용자에게 대한 특성정보에 의해 유사도  $S_R$ 를 계산한다. 이러한 일련의 과정을 통해 계산된 유사도를 평가한다. 사례를 선정하는 과정은 3.6절의 방법에 의해 이루어진다.

3.6 사례선정 알고리즘

구매가 고객 본인을 위한 경우와 타인을 위한 경우로 하여, 다음의 과정을 통해 사례가 선정되도록 한다.

- $S_{IG}$  : 구매고객 본인과 i 사례와의 유사도 값
- $S_{IR}$  : 상품의 실 사용자와 i 사례와의 유사도 값
- L : 타인을 위한 구매의 사례베이스의 크기
- P : 본인을 위한 구매의 사례베이스의 크기

```

max = 0 ;
IF(타인을 위한 구매)
    FOR(i=1 ; i ≤ L ; i++){
        i번 사례에 대해  $S_{IG}$ 와  $S_{IR}$ 을 각각 계산
        IF( $S_{IG} \geq 0.8, S_{IR} \geq 0.85, S_{IG} + S_{IR} \geq 1.7, \text{AND}$ 
            max <  $S_{IG} + S_{IR}$ )
            THEN max =  $S_{IG} + S_{IR}$  ;
    }
THEN max 값에 해당하는 사례를 선택한다.
ELSE IF(본인을 위한 구매)
    FOR(i = 1 ; i ≤ P ; i++){
        i번 사례에 대해  $S_{IG}$ 를 계산
        IF( $S_{IG} \geq 0.8 \text{ AND } \text{max} < S_{IG}$ )
            THEN max =  $S_{IG}$  ;
    }
THEN max 값에 해당하는 사례를 선택한다.
    
```

위의 사례선정 알고리즘은 타인을 위한 구매인 경우에 구매하는 사람보다 구매된 제품을 사용하는 사람의 특성을 우선시할 수 있도록 한다.

#### 4. 시스템 운용 및 분석

##### 4.1 시스템 운용

제한한 시스템을 구축하기 위해 O/S는 Windows 2000 Server, 데이터베이스 서버는 MS-SQL, 웹 서버는 IIS, 프로그래머는 ASP를 사용하였다. 본 연구에서 제시한 추천시스템을 테스트하기 위하여 실제 운영되고 있는 가방쇼핑몰의 판매정보와 고객정보를 이용하였다. 제품추천을 위한 시스템의 실행절차는 다음과 같다.

##### 4.1.1 구매유형 결정

고객이 제품을 구입할 목적으로 쇼핑 할 경우 구매유형을 결정하고 구매유형에 따라 제품을 추천해주는 실행과정이 선택되면 규칙베이스 검색을 하게 된다.

##### 4.1.2 회원고객을 위한 규칙베이스 검색

쇼핑몰에 접속한 고객의 최근구매 정보를 <표 4>과 같은 규칙베이스에서 검색해서 추천을 해준다. 규칙베이스는 판매기록을 바탕으로 Apriori 알고리즘에 의한 연관관계규칙에 의해 생성된 것이다.

<표 4> Rule\_base

num	rec_item	buy_item	total_buy	buy_rate	co_rate
1	AB_18	AB_21	32	0.32	0.688
2	AB_18	AB_38	28	0.28	0.714
3	AB_18	AB_23	18	0.18	0.778
4	AB_18	AB_40	14	0.14	0.714
5	AB_18	AB_29	12	0.12	0.667
6	AB_18	AB_22	10	0.1	0.6
7	AB_26	AB_19	10	0.1	0.8
∴	∴	∴	∴	∴	∴

여기에 해당되는 규칙을 발견하면 해당하는 제품을 추천하지만 규칙을 발견하지 못하거나 추가적으로 제품을 추천하기 위해서 사례베이스를 검색하게 된다.

##### 4.1.3 회원고객을 위한 사례베이스 검색

회원고객의 특성과 사례베이스-1간의 유사도를 검색해서 유사도 값이 임계점 이상인 사례의 제품을 고객에게 추천해준다. 그리고 이 단계에서 다른 사람을 위한 구매를 추천받을 수 있도록 설계되어 있어 원하지 않을 경우 추천시스템을 종료도 가능하도록 하였다. 타인을 위한 구매목적인 경우에 대해서 알아본다.

##### 4.1.4 타인을 위한 구매의 정보 입력

타인을 위한 구매를 선택하게 되면 제품을 사용하게 될 사람과의 관계를 배우자, 연인, 형제, 동료, 선/후배, 직장상사, 스승, 자식 등의 관계를 선택한다.

위의 항목에서 연인항목을 선택했다면, 다음으로 (그림 5)와 같이 제품을 전해주고자 하는 대상자의 고객 특성을 입력한다.

(그림 5) 고객특성 입력창

입력된 정보에 의해 관계가 3이고 입력된 고객특성들에 의해 상태가 G-남자, R-여자인 사례가 <표 5>와 같이 검색되어진다.

##### 4.1.5 타인을 위한 구매의 사례베이스 검색

타인을 위한 구매의 경우 사례베이스는 <표 5>와 같이 구성되어 있다. 기존고객의 정보와 고객이 입력한 타인의 정보가 <표 6>과 같을 때 유사도 계산 및 사례선정을 위해 각 특성의 가중치는 식 (3)에 의해 각 특성별 유사도 값을 구한다.

유사도 계산 후 사례선정은 3.6절의 사례선정 알고리즘에 조건을 만족하는 사례는 <표 7>과 같다.

다음으로 선정된 사례의 상품을 (그림 6)과 같이 제시한다. 주어진 예는 알고리즘의 조건을 만족하는 사례가 있어 추천이 가능하지만 만족하는 사례를 발견하지 못할 경우는 상태 R의 입력된 특성 자료만으로 사례를 검색, 평가, 선정해서 제시하게 된다.

<표 5> 타인구매를 위한 사례베이스

사례번호	유형	관계	상태	나이	성별	관심분야	학력	결혼여부	직업	지역	구입제품	성공횟수
1	2	3	G	23	남	34	대제	미혼	3	서울	AB_25	10
			R	24	여	23	대졸	미혼	17	경기도		
2	2	3	G	29	남	33	대졸	미혼	9	서울	AB_19	8
			R	25	여	22	초대졸	미혼	14	서울		
3	2	3	G	32	남	1	대졸	미혼	7	경기도	AB_20	4
			R	26	여	7	대졸	미혼	5	경기도		
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴

<표 6> 새로운 사례의 고객특성

·	·	·	·	29	남	영화/비디오	대학원졸	미혼	사무직	서울
·	·	·	·	27	여	외국어	대학원	미혼	대학원생	서울

<표 7> 선정조건을 만족하는 사례

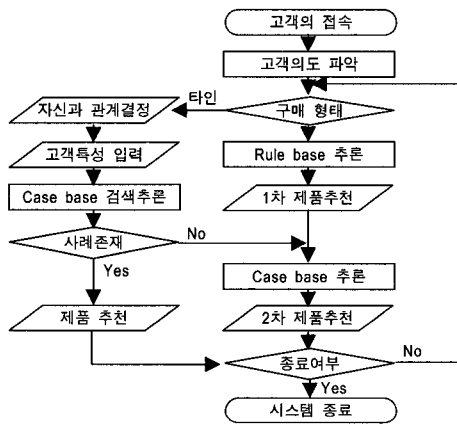
번호	유형	관계	상태	나이	성별	관심분야	학력	결혼여부	직업	지역	성공횟수	유사도값	구입제품
1	2	3	G	11.44	17	4.5	10.4	12	0	7	0.1	0.90	AB_25
			R	12.22	17	9	10.4	12	0	3.5	0.1	0.92	
2	2	3	G	13.00	17	9	10.4	12	0	7	0.08	0.96	AB_19
			R	12.48	17	4.5	7.8	12	0	7	0.08	0.86	
14	2	3	G	12.48	17	9	10.4	12	0	0	0.1	0.88	AB_56
			R	12.74	17	9	10.4	12	0	0	0.1	0.88	
20	2	3	G	11.96	17	4.5	13	12	0	7	0.08	0.92	AB_51
			R	13.00	17	0	13	12	7	7	0.08	0.96	

	여성용 브라다림단순타액 CT 127	27,000 원	클라이드	<input type="checkbox"/> 소량백의입기
	AB_19 [골프]클라이드	27,000 원	클라이드	<input type="checkbox"/> 소량백의입기
	정장용 슬리퍼 LMD 2432	45,200 원	아리모드	<input type="checkbox"/> 소량백의입기
	패션네그 수제 스타킹 트드백 LMD 2448	45,200 원	아리모드	<input type="checkbox"/> 소량백의입기

(그림 6) 최종 선정사례

4.1.6 사례 평가, 수정 및 등록 단계

자신을 위한 제품구매의 경우 추천한 제품을 구매하면 추천된 사례의 성공 횟수를 갱신하고 추천되지 않은 제품을 구입할 경우 고객의 특성과 구매된 제품으로 새로운 사례로 등록된다.



(그림 7) 제품추천의 흐름도

타인을 위한 구매의 경우 3.6절의 사례선정 알고리즘 추천된 제품을 구입할 경우 추천된 기존 사례베이스의 성공 횟수를 갱신하며, 추천된 제품이 아닌 다른 제품을 구입할 경우 구매자와 입력된 고객정보를 활용해서 사례베이스-2에 새로운 사례로 등록한다. (그림 7)은 고객의 구매의도에 따른 추천의 과정을 나타내고 있다.

5. 시스템 분석

CBR을 활용한 기존의 연구와 본 연구의 추천이 얼마나 고객에게 적절한지에 대해 조사하기 위해 실제 운영되고 있는 가방 쇼핑몰의 데이터를 이용하여 생성된 사례베이스를 바탕으로 설문문을 통해 수집한 10대~50대 100명의 고객을 대상으로 테스트를 하고 또 다른 5조의 자료는 난수발생에 의해 획득하여 가방에 대한 시험과 동일한 조건에서 테스트 한다. 난수에 의한 모든 자료는 균등분포(uniform distribution)를 따르는 것이고 데이터의 크기와 범위 또한 가방의 경우와 다소간 상이하다.

획득한 100명의 고객특성 중 절반은 자신을 위한 구매를 목적으로 나머지는 타인을 위한 구매를 가정하고 타인을 위

한 접속일 경우 타인의 정보를 추출하여 이를 적용한다.

쇼핑몰의 제품은 성별 구분이 있는 제품 28개와 구분이 없는 12개로 구성된다. 본인을 위한 구매의 경우 고객이 남자일 경우 남성용 제품이나 남녀 공용제품을 추천한 경우는 적절한 추천으로, 여성용 제품을 추천했을 경우는 적절하지 못한 추천으로 한다. 여자의 경우도 동일하게 적용한 결과 CBR을 이용한 기존 방법[13, 15]의 시험결과를 <표 8>에서 비교하고 있다. 자신을 위한 추천에서 부적절한 추천이 나오는 경우는 사례베이스가 충분하지 않은 경우에 해당 사례를 찾지 못한 경우이다. 또한 기존의 연구에서 부적절한 추천이 많은 것을 볼 수 있다. 아무리 사용자에게 대한 정보를 정확히 입력한다고 해도 실사용자와 구매자와의 관계를 확실히 하지 않으면 부적절한 추천이 될 가능성은 높다는 것을 의미한다. 남성에게 여성 제품을 여성에게 남성 제품을 추천하는 경우와 같이 적절하지 못한 추천을 한 경우이다. 기존의 연구에서는 총 평균 추천 성공률이 약 0.66이며, 본 연구의 성공률은 0.967로 나타나고 있다.

<표 8> 추천의 적절성 테스트

테스트	비고	사용자	추천 성공	추천 실패	총 추천수	성공율	평균 성공율
1	기존연구	자신	229	8	237	0.97	0.69
		타인	79	112	191	0.41	
	본 연구	자신	229	8	237	0.97	0.96
		타인	180	11	191	0.94	
2	기존연구	자신	230	6	236	0.97	0.64
		타인	60	125	195	0.31	
	본 연구	자신	230	6	236	0.97	0.95
		타인	181	14	195	0.93	
3	기존연구	자신	196	4	200	0.98	0.61
		타인	47	153	200	0.24	
	본 연구	자신	196	4	200	0.98	0.98
		타인	195	5	200	0.98	
4	기존연구	자신	239	11	250	0.97	0.64
		타인	78	172	250	0.31	
	본 연구	자신	242	8	250	0.97	0.97
		타인	243	7	250	0.97	
5	기존연구	자신	290	10	300	0.97	0.64
		타인	90	210	300	0.30	
	본 연구	자신	291	9	300	0.97	0.97
		타인	295	5	300	0.98	
6	기존연구	자신	331	19	350	0.95	0.71
		타인	162	188	350	0.46	
	본 연구	자신	335	15	350	0.96	0.97
		타인	341	9	350	0.97	

<표 9> 성별에 따른 추천의 적절성

구분	기존의 연구		본 연구	
	적절한 추천	부적절한 추천	적절한 추천	부적절한 추천
추천횟수	133	73	416	27
성공율	0.65		0.94	

<표 9>는 가방쇼핑몰의 자료를 이용하여 성별의 적합성이 고객의 특성에 부합하느냐를 고려하여 적절성을 평가한 것이다. 고객특성에 대한 고려는 타인을 위한 구매를 목적으로 접속한 고객에서 성별에 맞는 제품을 추천하였다더라도 제품사용자의 특성과 부합될 때 적절한 추천으로, 그렇지 않고 성별에 적합한 추천이 이루어졌지만 특성과 부합되지 않는 추천은 부적절한 추천으로 간주된다. 여기에서는 본 연구와 기존연구 결과의 적절성이 더욱 차이가 남을 보여준다. 결과적으로 기존의 제안된 방법[13, 15]에 비해 제안된 시스템의 적절성이 높게 나타나고 있으며 이는 상황별 사례로 고객의 선택 가능성이 높은 추천임을 보여주고 있다.

### 6. 결 론

e-쇼핑몰에서 고객위주의 개인화된 추천시스템을 적절히 활용함으로써 고객의 만족도와 신뢰도를 향상 할 수 있을 것이다. 이는 고객의 빈번한 방문을 유도하며 쇼핑몰의 인지도 및 경쟁력 제고를 통한 매출증대와 이익을 극대화 할 수 있는 것이다.

추천시스템은 고객의 의도를 정확한 파악하여 고객이 원하는 제품을 추천하여야 한다.

RBR과 CBR의 성공의 관건은 첫째 RBR을 위한 정확한 규칙의 발굴이며, 둘째 CBR을 위해 풍부한 사례의 축적이며 그리고 사례를 정확하게 평가하는 유사도 평가는 아무리 강조해도 지나치지 않다고 할 수 있을 것이다.

### 참 고 문 헌

[1] D. Kim and S. Kim, "Dynamic Expert Group Models for Recommender Systems," Proceedings of the First Asia-Pacific Conference on Web Intelligence: Research and Development, Maebashi City, Japan, October, 2001.

[2] A. Cliff, D. Kania and Y. Beth, "Internet World Guide to One-to-One Web Marketing," John Wiley & Sons, Inc., New York, 1998.

[3] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In Proceedings of the 20th VLDB conference, Santiago, Chile, Sept., 1994.

[4] "Clementine Data mining system ver. 6.5," SPSS Inc. Chicago, IL, USA.

[5] Aamodt, A. and E. Plaza, "Case-based reasoning: foundational issues, methodological variations, and system approaches," Artificial Intelligence Communications, Vol.7, No.1, 1994.

[6] Ramakrishnan Srikant and Rakesh Agrawal. "Mining generalized association rules," In 21st Int'l Conf. on Very Large Databases(VLDB), Zurich, Switzerland, Sept., 1995.

[7] R. Agrawal, T. Imielinski and A. Swami. "Mining association rules between sets of items in large databases," In Proc.

of the ACM SIGMOD Conference on Management of Data, Washington, D.C., May, 1993.

[8] 한정기, 주정애, 윤현준, "개인화란 무엇인가?", Onbit times, 2002.

[9] 현우석, 김용기, "선박에서 화재 탐재를 위한 규칙 및 사례기반 추론의 통합", 한국퍼지및지능시스템학회, 제10권 제1호, pp.303-306, 2000.

[10] 성백근, 김상희, 박덕원, "전자 상거래를 위한 사례기반추론의 판매지원 에이전트", 정보처리학회논문지D, 제7-D권 제5호, pp.1649-1656, 2001.

[11] 김은경, "신장 질환 진단을 위한 규칙 기반 추론과 사례기반추론의 통합", 정보과학논문지, 제24권 제10호, pp.1093-1100, 1997.

[12] 김진성, "데이터마이닝과 사례기반추론 기법에 기반한 인터넷 구매지원 시스템 구축에 관한 연구", 한국경영과학회지, 제28권 제3호, pp.135-148, 2003.

[13] 이경호, 윤창현, 박두순, "웹마이닝을 이용한 M-Commerce 추천시스템 설계 및 구현", 한국컴퓨터교육학회지, 제6권 제3호, pp.26-36, 2003.

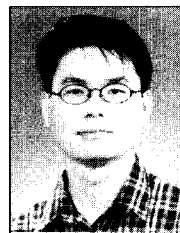
[14] 김영지, 문현정, 옥수호, 우용태, "사례기반추론 기법을 이용한 개인화된 추천시스템 설계 및 구현", 정보처리학회논문지 D, 제9-D권 제6호, pp.1009-1016, 2002.

[15] Finnie, G. and Sun, Z., "Similarity and Metrics in Case-Based Reasoning," International Journal of Intelligent Systems, Vol.17, pp.273-287, 2002.

[16] Fyfe, C. and Corchado, J. M., "Automating the Construction of CBR Systems using Kernel Methods," Journal of Intelligent Systems, Vol.16, pp.571-586, 2001.

[17] 허철희, 조성진, 정환목, "사례기반추론 에이전트를 이용한 전자상거래", 한국전자상거래학회, 제5권 제22호, pp.49-60, 2000.

[18] 허문렬, "수리통계학", 경문사, pp.53-54, 1989.



### 이 건 호

e-mail : ghlee@ssu.ac.kr  
 1986년 대구대학교 산업공학과(학사)  
 1991년 인하대학교 산업공학과(석사)  
 1996년 U. of Iowa, Dept. of Industrial Eng. (ph.D)  
 1997년~1999년 숭실대학교 산업공학과  
 전임강사

1999년~2004년 숭실대학교 산업정보시스템공학과 조교수  
 2004년~현재 숭실대학교 산업정보시스템공학과 부교수  
 관심분야 : 소프트웨어공학, E-비즈니스, 인공지능, 민첩생산공학



### 이 동 훈

e-mail : dhlee94@lycos.co.kr  
 2002년 대구대학교 산업공학과(학사)  
 2004년 숭실대학교 산업·정보시스템공학과  
 (석사)  
 관심분야 : E-비즈니스, 품질관리, 정보시스템