

인터넷 상점에서 개인화 광고를 위한 장바구니 분석 기법의 활용*

김종우** · 이경미***

Application of Market Basket Analysis to Personalized Advertisements on Internet Storefront*

Jong-Woo Kim** · Kyung-Mi Lee***

■ Abstract ■

Customization and personalization services are considered as a critical success factor to be a successful Internet store or web service provider. As a representative personalization technique, personalized recommendation techniques are studied and commercialized to suggest products or services to a customer of Internet storefronts based on demographics of the customer or based on an analysis of the past purchasing behavior of the customer. The underlining theories of recommendation techniques are statistics, data mining, artificial intelligence, and/or rule-based matching. In the rule-based approach for personalized recommendation, marketing rules for personalization are usually collected from marketing experts and are used to inference with customers data. However, it is difficult to extract marketing rules from marketing experts, and also difficult to validate and to maintain the constructed knowledge base. In this paper, we propose a marketing rule extraction technique for personalized recommendation on Internet storefronts using market basket analysis technique, a well-known data mining technique. Using marketing basket analysis technique, marketing rules for cross sales are extracted, and are used to provide personalized advertisement selection when a customer visits in an Internet store. An experiment has been performed to evaluate the effectiveness of proposed approach comparing with preference scoring approach and random selection.

* 본 연구는 일부 한국과학재단 Successful Post-doctoral Fellow 프로그램의 지원으로 이루어졌음.

** 충남대학교 통계학과 조교수

*** 충남대학교 통계학과

1. 서론

인터넷의 발전에 힘입어 글로벌 전자 경제로의 변혁이 빠른 속도로 진행되고 있다. 전자 상거래의 한 부분으로 기업-소비자간 전자상거래(business-to-consumer electronic commerce)가 다양한 제품 부문에서 증가하고 있다. 1997년부터 2001년 사이에 재무 서비스의 경우는, 12억불에서 50억불로, 온라인 여행 서비스의 경우는 8억6천3백만불에서 38억불로, 서적 및 음반 판매의 경우는 1억5천6백만불에서 11억불로 증가할 것으로 예상되고 있다 [10]. 이와 같이 증가되고 있는 기업-소비자간 온라인 시장에서 승자가 되기 위해서 인터넷 상점들은 글로벌 경쟁 속에서 경쟁 우위 확보를 위해 계속적으로 노력하고 있다.

90년대에 소개된 일대일 마케팅(one-to-one marketing)은 전통적인 대중 마케팅과 상품 중심의 마케팅에서 고객 중심의 마케팅으로의 근본적인 마케팅 기저의 변화를 제시하고 있다[21]. 일대일 마케팅은 관심의 초점을 시장점유율에 기반한 지역적 시장 점유에서 개별 고객의 점유로 옮겨놓았다. 개별 고객과의 지속적인 관계를 유지하기 위해서는 기술적으로는 고객 데이터베이스의 구축과 함께 쌍방향 의사소통 채널의 확보가 필요하다[4].

인터넷 상점 또는 웹 기반 정보 제공자는 실제로 일대일 마케팅을 수행할 수 있는 좋은 정보 시스템 환경을 확보하고 있다. 이러한 환경을 바탕으로 일대일 마케팅을 실현하는 도구로써 개인화 추천 기술들이 많이 제시되고 있으며, 실제로 Amazon, Yahoo, Reel, eToy, CDNOW, Moviefinder 등 유명 웹사이트들에서 적용되고 있다[4, 5, 9, 20]. 이러한 개인화 추천 기술은 고객 측면에서는 원하는 상품을 찾기 위한 수고와 시간을 줄여주고, 인터넷 상점 입장에서는 고객의 충성도(loyalty) 증대, 판매 증대, 광고의 수익성 증대, 타겟 홍보의 이익을 가져다 준다[5]. 개인화 추천 기술은 고객의 신상 정보와 인터넷 상점 내에서의 구매 행위에 대한 정보를 바탕으로 고객 취향을 반영한 상품이나 서

비스를 추천한다. 이러한 상품 추천 시스템에는 Net Perceptions사의 GroupLens Recommendation Engine, BroadVision사의 One-to-One Enterprise, LikeMinds사의 Preference Server 등이 있다[8, 19, 23].

현재 상용화된 추천 시스템들은 주로 협동적 필터링(collaborative filtering)이나 규칙 기반(rule-based) 기법들을 활용하고 있다. 규칙 기반 접근법에서는 마케팅 전문가로부터 제시되는 마케팅 규칙이 개인화 광고 선정에 핵심적인 역할을 한다. 따라서 규칙 기반 접근법의 효과성은 규칙 베이스에 저장되어 있는 마케팅 규칙의 품질에 전적으로 의존한다. 하지만 실제로 유용한 마케팅 규칙을 마케팅 전문가로부터 획득하는 것은 쉽지 않으며, 또한 구축된 규칙 베이스를 계속적으로 갱신하고 유지하는 것도 쉽지 않다.

최근에 경영과학 분야에서 활발히 적용이 모색되고 있는 데이터 마이닝은 대용량의 데이터로부터 의미있는 패턴이나 규칙을 발견하기 위한 자동적인, 또는 반자동적인 분석이나 탐색을 의미한다 [7]. 데이터 마이닝 기술들은 주로 통계학, 인공지능 기술에 기반을 하며, 주요한 데이터 마이닝 기술에는 메모리 기반 추론(memory-based reasoning), 의사결정 나무(decision tree), 장바구니 분석(market basket analysis), 자동 군집 분석(automatic cluster detection), 연결 분석(link analysis), 신경망(artificial neural networks), 유전자 알고리즘(genetic algorithm) 등이 있다. 본 연구는 데이터 마이닝 기법이 규칙 기반 추천 기법의 지식 추출 및 유지 문제에 유용하게 활용될 수 있다는 아이디어에서 출발되었다. 특히, 본 논문에서는 연관성 규칙을 추출하는데 주로 활용되는 장바구니 분석 기법이 규칙 기반 접근법을 도와주는데 어떻게 사용될 수 있으며 얼마나 효과적인지를 검토하기로 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대한 검토로, 규칙 기반 개인화 추천 기술과 장바구니 분석 기법에 대하여 살펴본다. 3장에서는 장바구니 분석 기법이 어떻게 규칙 기반 추

천에 활용될 수 있는 지를 검토한다. 즉, 장바구니 분석 기법을 활용한 마케팅 규칙의 추출과 추출된 규칙을 활용한 개인화된 광고 선정 방안에 대하여 살펴본다. 4장에서는 제시된 접근법의 효과성을 분석하기 위한 실험 결과를 제시한다. 5장에서는 결론을 제시한다.

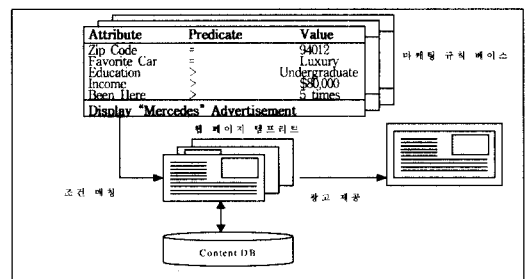
2. 관련 연구

2.1 규칙기반 개인화 추천 기술

현재 활용 또는 연구되고 있는 개인화 추천 기술로는 협동적 필터링, 규칙 기반 기법, 선호도 점수법 등이 있다. 협동적 필터링 기법은 고객들의 광고에 대한 평가 정보들을 기반으로 하여, 유사한 선호도를 갖고 있는 고객들의 평가를 활용하여 개인화된 광고를 제시한다. 협동적 필터링 기법에 대한 주요 연구에는 미네소타 대학의 GroupLens, 버클리 대학의 Jester 등이 있으며, 주요 상용화 제품에는 Net Perceptions사의 GroupLens Recommendation Engine, LikeMinds사의 Preference Server, WiseWire사의 WiseWire.com, Firefly Network사의 Firefly Passport Office 등이 있다[12, 14, 19, 20, 23]. BroadVision사에 의해서 처음으로 상용화되었던 규칙 기반 추천 기술은 인공지능 부문에서 활용되는 규칙 매칭 기법을 활용하여 적절한 제품을 추천하는데, 인터넷 라디오, 인터넷 텔레비전, 인터넷 은행, 투자관리 회사의 ICRM(Internet Customer Relationship Management) 등에 적용되고 있다[8, 9, 15]. 규칙 매칭 기법을 활용한 상용화된 상품 추천 시스템에는 Art Technology Group사의 Dynamo Relationship Commerce Suite, BroadVision사의 One-to-One Enterprise, GuestTrack사의 GuestTrack 등이 있다[6, 9, 13]. 규칙 기반 추천 시스템들은 특히 은행, 투자관리 회사 등 금융 회사 사이트에서 ICRM 시스템의 일부로 활발히 활용되고 있다. 예를 들어, Fidelity Investments, John Hancock Funds, Oppenheimer Funds, Scud-

der Investments, Liberty Financial Companies 등에서 Art Technology Group사나 BroadVision사의 제품을 활용하여 ICRM을 구축하여 서비스하고 있다[15]. 협동적 필터링 기법이나 규칙 기반 기법과는 다른, 자기 학습(self learning) 기반 기법의 하나로 선호도 점수법이 제시되었다[2]. 이 기법에서는 고객의 초기 프로파일과 구매 이력, 인터넷 상점 내에서의 행위 정보를 바탕으로 고객 선호도에 적합한 광고 선정을 피하고 있다. 즉, 선호도 점수 개념을 도입하여 고객의 제품군별 선호도 점수를 고객의 상점 내 행위를 바탕으로 계속적으로 갱신하고, 이 점수를 활용하여 개인화된 광고를 선정한다.

규칙 기반 접근법에서는 규칙 베이스의 마케팅 규칙이 개인화 추천의 핵심적인 역할을 한다<그림 1 참조>. 예를 들어, “서울에 사는 10대 소년들은 농구를 좋아한다”는 규칙을 기초로 하여, 인터넷 상점에 접속한 고객이 서울에 사는 10대 소년인 경우에는 농구에 관련된 뉴스나 농구화에 대한 광고를 제시하게 된다. 이러한 마케팅 규칙들은 일반적으로 해당 분야의 마케팅 전문가에 의해서 제시된다. 하지만 이와 같은 마케팅 규칙의 추출은 많은 경우 마케팅 전문가의 직관에 의존하는 경우가 많으며, 규칙 추출 프로세스도 많은 시간과 노력을 필요로 한다. 또한 이러한 마케팅 규칙들이 계속적으로 추가되고 갱신되어야 하는데, 이러한 작업들도 전문가의 직관이나 지식에 계속적으로 의존하게 된다.



<그림 1> 개인화 추천을 위한 규칙 기반 접근법 (Source : [8])

2.2 장바구니 분석 기법

장바구니 분석이라고 불리우는 자동화된 연관성 규칙 추출 기술은 교차 판매(cross sales)에 활용되는 대표적인 데이터 마이닝 기법이다. 즉, 장바구니 분석을 통해서 X 라는 상품이 Y 라는 상품과 함께 팔리는 확률이 높다는 연관성 규칙을 추출하고, 이를 상품 진열 레이아웃, 상품 패키징 등에 활용한다[1, 7, 25].

이러한 연관성 규칙을 추출할 때는 몇 가지 수량화 된 기준들이 활용된다. 예를 들어, X 라는 상품을 구매하는 사람의 40%는 Y 라는 상품을 구매하며, 이와 같이 X, Y 상품을 함께 구매하는 경우가 전체 거래의 5%에 해당한다.와 같이 연관성 규칙의 유용성을 판단하는 수치들이 함께 제시되는 것이 필요하다. 연관성 규칙을 수량화 하는 기준으로 *Support*, *Confidence*, *Lift* 세가지가 있다. 만일 관심 있는 규칙이 'X라는 상품을 구입한 사람은 Y라는 상품도 구입한다.'라고 하면, 연관성 규칙을 평가하는 세 가지 기준은 다음과 같다.

- (1) *Support*는 전체 거래 중 X 와 Y 를 포함하는 거래량이 어느 정도인가를 파악하는 측정의 기준이고, 다음과 같이 확률로 나타낸다.

$$Support = \frac{n(X \cap Y)}{N}$$

이때, N 은 전체 거래의 회수이고, $n(X \cap Y)$ 는 X 라는 상품과 Y 라는 상품을 함께 구매한 빈도수이다.

- (2) *Confidence*는 X 라는 상품을 구매한 거래 중에서, Y 라는 상품이 포함된 거래의 정도를 측정하는 기준이다. 이는 연관성 규칙의 강도를 나타내며, 다음과 같이 조건부확률로 나타낸다.

$$Confidence = P(Y | X)$$

- (3) *Lift*는 임의로 Y 를 구매하는 경우에 비해, X

와의 관계가 고려되어 구매되는 경우의 비를 측정하는 기준이다. *Lift*는 확률로 나타내지 않으며, 다음과 같이 나타낼 수 있다.

$$Lift = \frac{P(Y | X)}{P(Y)} = \frac{P(X \cap Y)}{P(X) \cdot P(Y)}$$

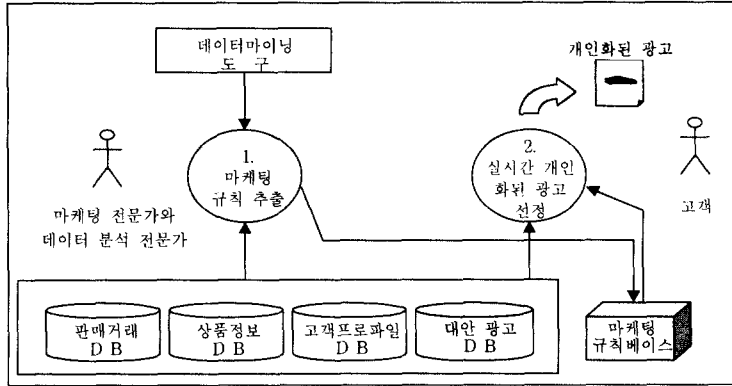
이때, $P(Y)$ 는 전체 거래 중에서 Y 라는 상품의 거래가 일어나는 확률이다.

*Lift*의 값이 1에 가깝다는 것은 통계적으로 서로 독립이라는 것을 뜻하게 되고, 1보다 크다는 것은 양의 연관관계를, 1보다 작으면 음의 연관관계를 갖는다는 것을 의미한다. 장바구니 분석에서는 위의 3가지 기준들을 사용하여 연관성 규칙을 추출한다. 의미 있는 연관성 규칙이 되기 위해서는 *Confidence*의 값이 크면 좋은 결과이지만, *Support*도 어느 정도 수준에 도달하여야 한다. 또한, *Confidence*와 *Support*는 자주 구매 되는 항목에 대하여 높게 발생 할 수 있다. 이런 경우, *Lift*를 함께 고려하여야 한다[1, 25]. 어떤 연관성 규칙의 유용성 여부를 판단하기 위한 *Confidence*나 *Support*에 대한 절대적인 기준값이 이론적으로 결정되어 있는 것은 아니며, 분석자의 경험과 판단에 의하여 반복적으로 값을 변경하며 의미 있는 규칙을 찾아내야 한다. 연관성 규칙을 데이터베이스의 데이터로부터 효율적으로 찾아주는 알고리즘으로는 IBM 연구소에서 연구된 Apriori 알고리즘과 그 변형(AprioriTid, AprioriHybrid)들이 대표적이며, 이외에도 Houtsma와 Swami의 알고리즘, Savasere 등이 제시한 알고리즘 등이 존재한다[3, 16, 22, 24].

3. 마케팅 규칙의 추출과 개인화 광고 선정

3.1 데이터 마이닝 기법을 활용한 마케팅 규칙의 선정

규칙 기반 접근법에 데이터 마이닝 기술을 활용하기 위해서 <그림 2>와 같이 2단계 접근법을 제



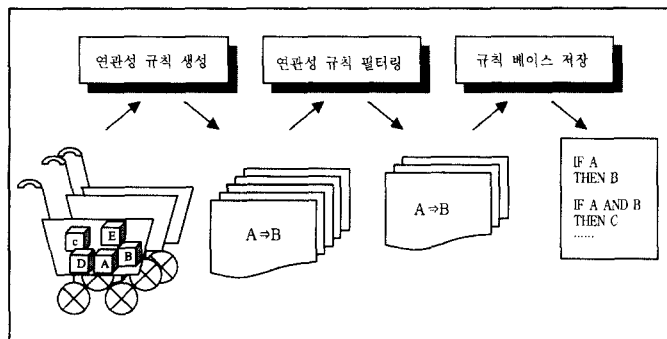
<그림 2> 인터넷 상점에서 개인화 광고를 위한 2단계 모형

시한다. 첫번째 단계는 마케팅 규칙 추출 단계로 마케팅 전문가, 데이터 분석 전문가들이 배치 모드로 데이터 마이닝 도구를 사용하여 마케팅 규칙을 추출하는 단계이다. 의미있는 마케팅 규칙을 추출하기 위해서 판매 거래 데이터베이스, 고객 프로파일 데이터베이스, 상품정보 데이터베이스 등이 활용된다. 추출된 마케팅 규칙은 마케팅 규칙 베이스에 저장이 되고, 이는 고객이 인터넷 상점을 방문 시에 실시간 개인화 광고 선정을 위해서 활용된다. 즉, 고객이 인터넷 상점을 방문 시에 저장되어 있는 마케팅 규칙들을 바탕으로 대안 광고 데이터베이스의 대안 광고들 중에서 고객의 취향에 맞는 광고들을 선택하여 고객에게 제공한다.

3.2 마케팅 규칙의 추출

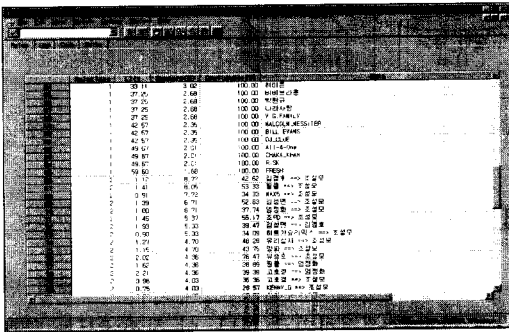
설명을 위해서 음악 CD를 판매하는 가상의 인터넷 상점을 설정하기로 한다. 이 상점에서는 클래식, 대중가요, 모음곡이나 영화음악 CD 등을 판매하고, 고객은 인터넷 상점에 최초 방문 시에 개인의 일반적인 인상 정보과 함께 선호하는 제품군을 선택한다고 가정한다.

장바구니 분석을 이용한 마케팅 규칙의 추출은 <그림 3>과 같이 3 단계; (1) 연관성 규칙의 생성, (2) 연관성 규칙의 필터링, (3) 규칙 베이스에 저장의 순서로 수행된다. 연관성 규칙 생성 단계에서는 개별 고객이 지금까지 구매한 상품들을 하나의 장바구니로 보아서 판매 거래 데이터에 대하여



<그림 3> 마케팅 규칙의 추출

장바구니 분석을 행한다. 장바구니 분석은 데이터 마이닝 도구의 장바구니 분석 알고리즘을 사용하여 수행되며, 결과로 연관성 규칙들이 생성된다. 예를 들어 <그림 4>는 데이터마이닝 도구인 SAS Enterprise Miner의 장바구니 분석 기능을 활용하여 실제 연관성 규칙을 생성한 결과 화면이다[1].



<그림 4> SAS Enterprise Miner를 활용한 연관성 규칙 생성 결과 화면

다음 단계는 생성된 연관성 규칙 중에 개인화 광고에 활용할 연관성 규칙을 필터링하는 단계이다. 이때 *Support*, *Confidence*, *Lift* 값을 기준으로 하여 유용한 연관성 규칙을 필터링한다. 2.2절에서 언급하였듯이, 연관성 규칙의 채택 여부를 결정하기 위한 고정된 최적의 *Support*, *Confidence*, *Lift* 값이 존재하는 것은 아니며, 2.2절에서 제시된 원칙들을 바탕으로 분석자의 경험과 판단에 따라서 반복적으로 세 개 기준값을 변경하면서 결정한다. 예를 들어, 규칙을 필터링하는 기준 *Confidence* 값을 높이면, 생성되는 규칙의 수는 줄어드나, 연관성의 강도가 큰 규칙들만이 필터링된다. 따라서, *Support*, *Confidence*, *Lift* 기준값이 높아지면, 필터링된 규칙들의 정확성이나 유용성은 높아지나 규칙의 수가 적어지는 trade-off가 존재한다. <표 1>은 <그림 4>에서 생성된 연관성 규칙 중에서 필터링 후의 연관성 규칙의 일부를 보여준다.

필터링된 연관성 규칙들은 추후에 개인화된 광고 선정에 활용할 수 있는 형태로 변환되어야 한다. 필터링된 규칙은 추론 엔진이 처리가능한 형태

<표 1> 필터링 후의 연관성 규칙의 예

CONF (%)	LIFT	SUPP (%)	RULE
50.00	7.10	2.01	ASHKENASY ==> ANNE_SOPHIE_MUTTE
46.15	3.06	2.01	조성모 & 유승준 ==> 핑클
45.00	2.20	3.02	조성모 & 김성면 ==> 김경호
42.86	2.84	2.01	조성모 & 양파 ==> 핑클
40.00	2.71	3.36	CLUB_DJ_가요리믹스 ==> 히트가요리믹스
39.47	1.93	5.03	김성면 ==> 김경호
35.29	2.34	2.01	유승준 ==> 핑클
35.00	2.37	2.35	약속 ==> 히트가요리믹스
31.25	1.76	3.36	양파 ==> 김정화
30.00	2.79	2.01	약속 ==> 쉬리
27.59	1.87	2.68	조PD ==> 히트가요리믹스
27.27	3.87	2.01	정트리오 ==> ANNE_SOPHIE_MUTTE
27.27	1.53	2.01	정트리오 ==> 김정화
26.09	2.43	2.01	조성모 & MAX5 ==> 쉬리
25.00	1.66	2.68	양파 ==> 핑클

인 'IF...THEN' 규칙 형태로 변환되어 저장된다. 저장되는 연관성 규칙들은 조건부와 결론부뿐만 아니라, *Confidence* 값을 함께 저장하도록 한다. 이 *Confidence* 값은 개인화 광고 선정을 위한 추론 시에 동시에 조건부를 만족하는 규칙이 두 개 이상인 경우, 우선순위를 부여하기 위해서 활용된다. *Confidence* 값을 이러한 목적으로 활용하는 것은 *Confidence* 값이 조건부 확률값($P(Y|X)$)으로 연관성의 강도를 나타내기 때문이다. 변환된 연관성 규칙의 구조는 다음과 같다.

<rule> :=	'RULE' <rule_id> 'CONFIDENCE' confidence 'IF' condition 'THEN' conclusion
confidence :=	<REAL>
condition :=	purchase ('AND' purchase)*
conclusion :=	'display('<product_id>')
purchase :=	'purchase('<product_id>')
Note.	:= definition of nonterminal token

'...' reserved token
 <...> value-bearing
 {...}*zero or more iterations of contents limited
 by the braces

예를 들어, <표 1>에 연관성 규칙 중 처음 두 개의 규칙을 규칙 베이스에 저장하기 위한 형태로 변환하면 다음과 같다.

```

RULE ass_rule01
CONFIDENCE 50.00
IF purchase(askenasy)
THEN display(anne_sophie_mutte)

RULE ass_rule02
CONFIDENCE 46.15
IF purchase(조성모) AND purchase(유승준)
THEN display(핑크)
    
```

3.3 광고의 선정

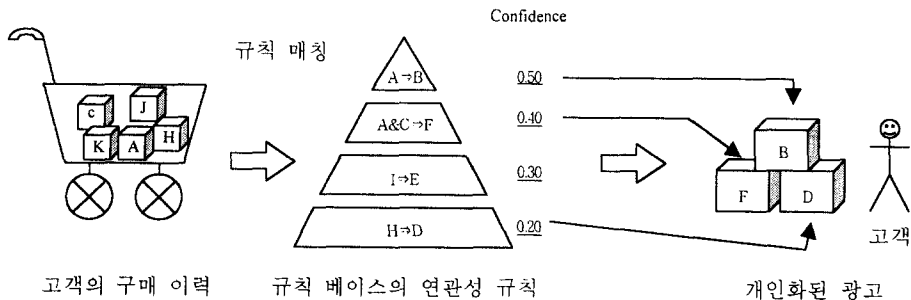
고객이 인터넷 상점을 방문 시에 해당 고객의 과거 구매 이력과 저장된 연관성 규칙을 매칭시켜서 개인화된 광고를 선정한다. 즉, <그림 5>에서와 같이 과거에 A, C, H, J, K 상품을 구매한 고객이 상점을 방문한 경우, 마케팅 규칙 베이스의 연관성 규칙 중 우선순위가 높은 규칙부터 우선적으로 매칭이 되어, B, F, D 상품의 광고가 차례로 선정된다. 매칭되는 광고가 없는 경우에는 고객 등록 시 지정한 선호 제품군에 대한 정보를 바탕으로

로 광고를 선정한다. 만일 선정하고자 하는 광고의 수가 N인 경우, 보다 상세한 선정 알고리즘은 다음과 같다.

Selection Algorithm

- [STEP 1] 데이터베이스의 고객 프로파일, 구매 정보를 추론을 위한 사실(fact)로 변환
- [STEP 2] 구매 정보에 대한 사실들과 연관성 규칙을 매칭, 이 때 규칙 매칭의 우선 순위는 Confidence값을 기준으로 적용
- [STEP 3] 선정된 광고의 수가 N보다 작으면, 고객 등록 시 지정한 선호 제품군에 속한 상품 중 임의로 선정

<그림 6>은 광고 선정 알고리즘을 CLIPS(C Language Integrated Production System)을 활용하여 구현한 일부이다[11]. <그림 6A> 부분은 고객(customer), 해당 고객이 과거 구매한 상품에 대한 정보(purchase)와 고객에게 제공되도록 선정된 광고들에 대한 사실들(display)을 저장하기 위한 템플릿들의 정의 부분이다. <그림 6B> 부분은 해당 고객의 고유번호와 선호 제품군에 대한 사실로, 고객 프로파일 데이터베이스로부터 추출되어서 지식 베이스에 사실(fact)로 저장된다. <그림 6C>는 해당 고객이 과거에 구매한 상품들에 대한 정보로 판매거래 데이터베이스로부터 추출되어 사실로 저장되었다. <그림 6D>부분은 마케팅 규칙 추출 단



<그림 5> 연관성 규칙에 의한 광고 선정

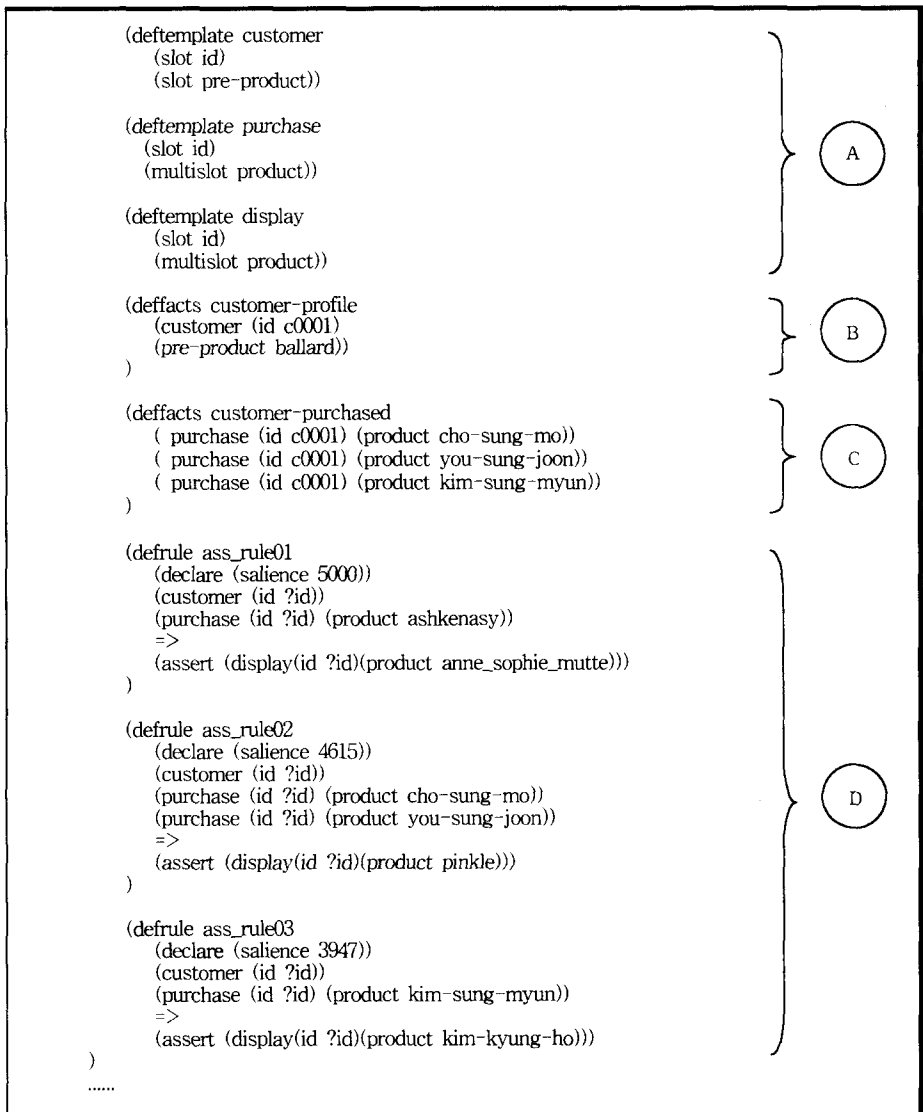
계에서 생성된 마케팅 규칙들로, 앞의 3.2 <표 1>의 연관성 규칙을 CLIPS에서 구현한 것의 일부이다. CLIPS에서는 규칙 매칭의 우선 순위를 부여하기 위하여 규칙의 특별한 속성으로 'saliency'를 사용한다[11]. 즉, 동시에 매칭가능한 규칙이 여러 개 있는 경우에 'saliency' 값이 높은 순서로 적용된다. 실제로 CLIPS에서 'saliency'값은 -100000에서 +100000까지의 정수값을 갖도록 되어있기 때문에

Confidence값의 100배 한 값을 저장하였다. 즉, 첫째 규칙의 'saliency'값 5000은 Confidence값 50.00이 매칭된 것이다.

4. 알고리즘 비교 실험

4.1 실험 내용

앞 절에서 제시한 개인화 광고 선정에 장바구니



<그림 6> CLIPS로 구현한 광고 선정 알고리즘의 일부

분석 기법의 활용에 대한 효과성을 평가하기 위하여 선호도 점수법과 임의 선정과 비교실험을 수행하였다. 실험은 두번에 걸쳐서 이루어졌는데, 첫번째 실험은 연관성 규칙을 생성하기 위해서 실시하였으며, 두번째 실험은 기법간의 효과성 비교를 위해서 실시하였다. 첫번째 실험에 대한 개요적인 내용은 다음과 같다.

- 조사대상 : 인터넷 리서치 회원
- 조사방법 : 인터넷 설문
- 표 본 수 : 208명
- 음악 CD 데이터 : 72개
 - 클래식(16개) : 현대, 고전
 - 대중가요(40개) : 발라드, 댄스, 랩, 락, R&B
 - 기타(16개) : 영화음악, 모음곡

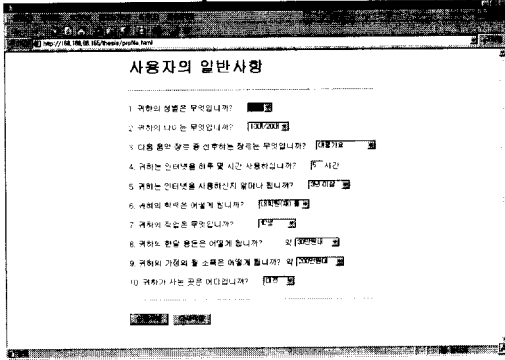
첫번째 실험에서는 응답자에게 72개의 음악 CD를 제시하고 이 가운데서 자신이 구매하고자 하는 CD를 5개 선택하도록 하였다. 이렇게 얻어진 구매 데이터를 이용하여 SAS Enterprise Miner를 이용하여 연관성 규칙을 생성하였다. 규칙을 필터링하기 위해 세가지 평가 기준은 $Lift = 1.3$, $Support = 2\%$, $Confidence = 25\%$ 로 지정하였다. 여기서 $Lift$, $Support$, $Confidence$ 의 기준값은 사전적으로 정해져 있었던 것은 아니며, $Lift$ 의 경우만 정의에 의해 1 이상으로 제약하고, 분석자가 기준값을 변화시키면서 규칙을 생성하여 적당한 수준에서 결정한 값이다. 실제로 <표 1>은 첫번째 실험 결과로 생성된 규칙의 일부이다.

두번째 실험은 세 가지 기법(연관성 규칙을 활용하는 방안, 선호도 점수법, 임의 선정법)간의 효과성 비교를 위한 실험으로 행해졌다. 실험에 대한 개요는 다음과 같다.

- 조사대상 : 인터넷 리서치 회원
- 조사방법 : 인터넷 설문

- 표 본 수 : 208명
- 구매 대상 음악 CD 데이터 : 36개(첫번째 실험에서 사용된 72개의 데이터 중 일부)
 - 클래식(8개) : 현대, 고전
 - 대중가요(20개) : 발라드, 댄스, 랩, 락, R&B
 - 기타(8개) : 영화음악, 모음곡
- 선호도 체크 대상 음악 CD 데이터 : 36개(첫번째 실험에서 사용된 72개의 데이터 중 구매대상을 제외한 나머지 CD 데이터)

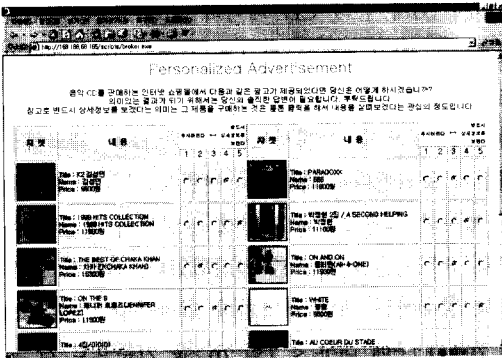
개인화된 광고 제공 기법의 효과를 살펴보기 위해 다음과 같은 절차로 실험을 수행하였다. (1) 첫번째 페이지에서 성별, 연령, 선호하는 음악장르 등 응답자의 일반사항을 입력한다<그림 7a>. (2) 두번째 페이지에서는 36개의 음악 CD 상품이 제시되고, 이 중 5장의 CD를 구매한다<그림 7b>. (3) 세번째 페이지에서는 세 가지 기법의 의해 선정된 광고가 각각 세 개씩 제시된다. 즉, 첫번째 실험에서 사용한 72개의 음악 CD 데이터 중 두번째 페이지 <그림 7b>에서 제시된 36개를 제외한 나머지 36개 음악 CD 데이터 중에서 각 기법을 활용하여 각각 3개씩 광고를 선정하였다. 응답자는 이 광고 각각에 대하여 5점 척도(1 : 무시하겠다 ~ 5 : 반드시 상세정보를 보겠다)를 체크하면 실험이 종료된다<그림 7c>. 연관성 규칙을 활용하여 제시되는 광고들은 첫번째 실험에서 생성된 마케팅 규칙들을 바탕으로 선정되었다. 선호도 점수법의 경우에는 해당 응답자의 일반사항 정보 <그림 7a>와 구매 내용 <그림 7b>을 기준으로 각 제품군별 선호도 점수를 부여하여, 이 점수가 높은 제품군의 광고를 우선적으로 제공하였다(선호도 점수법에 대한 보다 상세한 내용은 [2]참조). 임의 선정의 경우에는 난수를 발생시켜서 대안 광고 중에서 임의로 광고를 선정하여 제공하였다.



(a) 응답자 일반사항 입력



(b) 상품 구매 : 상품 36개 중 원하는 음악 CD 5장을 구입



(c) 광고 선호도 체크 : 연관성 규칙, 선호도 점수, 임의 선정에 의해 광고가 각각 3개씩 선정되어 함께 제공되며, 응답자는 제공된 광고에 대한 관심의 정도를 5점 척도로 체크한다.

<그림 7> 실험 수행 절차

4.2 실험 결과

각 기법별로 각 고객에게 제시된 3개의 광고에 대한 5점 척도의 평균값(최소값은 1, 최대값은 5)을 비교하였다. 두번째 실험에서 208명의 응답 결과 각 기법의 평균값은 <표 2>와 같다. <표 2>를 보면, 연관성 규칙을 사용한 경우가 가장 높고, 선호도 점수법이 차순임을 알 수 있다. 이들 평균값들간의 통계적으로 유의한 차이가 있는지를 살펴 보기 위해 반복측정 분산 분석법(Repeated Measures ANOVA)을 실시하였다[17]. 반복측정 분산 분석법을 사용한 이유는, 동일한 고객에 대하여 여러 기법으로 제시된 광고를 함께 제공하였기 때문이다. 반복측정 분산 분석의 결과는 <표 3>, <표 4>와 같다.

<표 2> 기법들의 평균

기법	평균
연관성 규칙	3.27
선호도 점수	3.13
임의 선정	2.72

<표 3> 반복측정 분산 분석

반복 변수	F	Pr > F
임의선정/선호도 점수/연관성 규칙	20.2783	0.0001

<표 4> Contrast Variables 분석

Contrast Variable	F	Pr > F
임의선정-선호도 점수	14.59	0.0002
임의선정-연관성 규칙	40.51	0.0001
선호도 점수-연관성 규칙	1.62	0.2048

분산분석 결과 <표 3>에서와 같이 5% 유의수준에서 세가지 기법간의 통계적으로 유의한 차이가 있음을 알 수 있었다. 각각에 대해 Contrast 분석 결과, 연관성 규칙을 사용한 경우와 임의 선정과는 5% 유의 수준에서 통계적으로 유의한 차이가 있으나, 선호도 점수법과 연관성 규칙을 사용한

경우간에는 유의한 차이가 없음을 알 수 있었다 (<표 4> 참조). 비록 평균값은 연관성 규칙을 사용한 경우가 선호도 점수법보다 높게 나타났지만, 통계적으로는 큰 차이가 없다고 여겨진다. 이에 대해서는 추후에 추가적인 분석(예를 들어, 타 상품군을 기준으로 한 비교)이 필요할 것으로 생각된다. 또한, 고객 특성 정보(나이, 성별 등)와 결합된 연관성 규칙의 생성과 효과성 검토에 대한 실험도 추가적으로 필요하다.

5. 결 론

본 논문에서는 인터넷 상점에서의 일대일 마케팅의 일환으로 개인화 광고를 제공하기 위한 데이터 마이닝 기법의 응용, 특히 장바구니 분석 기법의 응용 방안을 제시하였다. 장바구니 분석 기법을 사용하여 교차 판매를 위한 상품 판매간의 연관성 규칙들을 도출하고, 이들 연관성 규칙과 고객의 과거 구매 데이터를 활용하여 개인화된 상품 추천이 가능하다. 본 논문에서는 연관성 규칙의 도출을 위한 단계로, (1) 연관성 규칙의 생성, (2) 연관성 규칙의 필터링, (3) 규칙 베이스에 지장을 제시하였다. 또한 이들 연관성 규칙을 활용한 개인화 상품 추천 알고리즘과 프로토타이핑을 제시하였다. 제시된 기법의 유용성을 검토하기 위하여, 임의로 광고를 선정할 경우, 선호도 점수법, 연관성 규칙을 활용하는 방법에 대한 효과성을 비교 분석하였다. 비교 분석 결과 연관성 규칙을 활용하는 경우가 가장 좋은 것으로 나타났으나, 통계적으로는 선호도 점수법과 연관성 규칙을 활용하는 방법의 차이가 유의하지는 않았다.

본 연구에서 제시한 장바구니 분석 기법 이외에도 다양한 데이터 마이닝 기법들이 인터넷 상점에서 개인화 광고 선정을 위해 활용될 수 있을 것으로 보인다. 실제로 인공지능 기술의 하나인 의사결정나무 추론 기법을 활용한 마케팅 규칙 추출에 대한 연구가 실행되었으며, 유용성을 갖는 것으로 나타났다[18]. 따라서 타 데이터 마이닝 기법에 대

한 활용 가능성에 대한 검토와 함께, 여러 가지 데이터 마이닝 기법들을 통합적으로 활용하고, 이들간의 효과성을 비교 분석하는 연구가 계속적으로 필요하다. 또한 선호도 점수법이나 협동적 필터링 기법과의 통합적 활용 방안에 대한 모색도 필요하다.

기업-소비자간 전자상거래(business-to-consumer electronic commerce)는 경영과학 기법들의 새로운 활용 분야가 될 것으로 보여진다. 경영과학의 대부분의 기법들은 데이터를 기반으로 한 과학적이고 합리적인 문제 해결 방안들이다. 인터넷 상점들은 인터넷 기반의 정보 시스템 인프라를 이미 보유하고 있으며, 이것은 경영과학 기법들을 활용할 수 있는 데이터를 이미 축적하고 있다는 것을 의미한다. 실제로 인터넷 상점들은 많은 부분에서 자동화된 지능적인 의사결정을 필요로 하고, 이들은 데이터베이스 기술과 접목된 경영과학 기법들의 활용을 요구한다. 이러한 측면에서 경영과학 기법, 특히 데이터 마이닝 기법과 자동 학습 기법들은 인터넷 상점의 개인화된 광고 제공, 지능형 고객 인터페이스 제공, 지능형 고객 서비스 창출, 마케팅 전략 수립 등에 유용하게 활용될 수 있다. 따라서 이를 위한 (1) 기업-소비자간 전자상거래를 위한 경영과학 기법들의 적용 가능성 모색, (2) 기업-소비자간 전자상거래 상황에 적합한 문제 해결 알고리즘의 개발, (3) 경영과학 기법의 활용 효과성 및 효율성의 검증 등에 대한 연구가 계속적으로 필요하다.

참 고 문 헌

- [1] 강현철, 한상태, 최종우, 김은석, 김미경, 「SAS Enterprise Miner를 활용한 데이터마이닝」, 자유아카데미, 1999.
- [2] 김종우, 이경미, 김영국, 유관중, “인터넷 상점에서의 실시간 개인화 광고 제공 기법”, 「경영정보학연구」, 제9권, 제4호(1999).
- [3] Agrawal, R. and R. Srikant, “Fast Algorithms for Mining Association Rules,” *Pro-*

- ceeding of the 20th Int'l Conference on Vary Large Databases, Santiago, Chile, Sept., (1994).
- [4] Allen, Cliff, Deborah Kania, and Yaeckel Beth, *Internet World Guide to One-to-One Web Marketing*, John Wiley & Sons, Inc., New York, 1998.
- [5] Ansari, Asim, Skander Essegaier, and Rajeev Kohli, "Internet Recommendation Systems," *Journal of Marketing Research*, Vol.XXXVII, August, (2000).
- [6] Art Technology Group, <http://www.atg.com>, (2000)
- [7] Berry, Michael J. A. and Gordon Linoff, *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley Computer Publishing, New York, 1997.
- [8] BroadVision, "A BroadVision One-to-One White Paper," *White Paper*, BroadVision, (1996).
- [9] BroadVision <http://broadvision.com>, (2000)
- [10] Business Week Graphics: On-line Sales Are Soaring, *Business Week*, June 22, (1998)(based on data from Forrester Research Inc.).
- [11] CLIPS, *CLIPS Reference Manual (Version 6.05)*, 1997.
- [12] Dragon, Richard V., "Recommendation Systems - Advice From the Web," *PC Magazin*, September 9, (1997).
- [13] GuestTrack Inc., <http://www.GuestTrack.com>, (2000)
- [14] Gupta, Ohruv, Mark Digiovanni, Hiro Norita, Ken Goldberg, "Jester 2.0 : Evaluation of a New Linear Time Collaborative Filtering Algorithm," *22nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, August, (1999).
- [15] Hegarty, Robert, "Cross-selling Fund Customers: Investment Managers Take CRM to the Internet," *Research Note*, TowerGroup, June, (2000).
- [16] Houtsma, M. A. W. and A. N. Swami, "Set-oriented Mining for Association Rules in Relational Databases," *Proceeding of the 11th International Conference on Data Engineering*, Taipei, Taiwan, (1995), pp.25-33.
- [17] Johnson, Richard A. and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall International, Inc., New Jersey, 1992.
- [18] Kim, Jong Woo, Byung Heon Lee, Michael J. Shaw, Hsin-Lu Chang, Matthew L. Nelson, "Application of Decision Tree Induction Techniques to Personalized Advertisements on Internet Storefront," submitted to *International Journal of Electronic Commerce*, (2000).
- [19] LikeMinds, "LikeMinds™ Personalization Server 3.0 Technology Backgrounder," *White Paper*, Andromedia, Inc., (1999).
- [20] Net Perceptions, <http://netperceptions.com>, (2000)
- [21] Peppers, Don and Martha Rogers, *The One to One Future : Building Relationships One Customer at a Time*, Currency and Doubleday, New York, 1993.
- [22] Savasere, A., E. Omiecinski, and Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases," *Proceeding of the 21th VLDB Conference*, Zurich, Switzerland, (1995). pp.432-443.
- [23] Schafer, J. Ben, Joseph Konstan, and John Riedl, "Recommendation System in E-Commerce," *Proceedings of the ACM Conference on Electronic Commerce*, November 3-5, (1999).
- [24] Srikant, R. and R. Agrawal, "Mining Generalized Association Rules," *Proceeding of the 21th VLDB Conference*, Zurich, Switzerland, (1995). pp.407-419.
- [25] Westphal, Christopher and Teresa Blaxton, *Data Mining Solutions : Methods and Tools for Solving Real-World Problems*, John Wiley & Sons, Inc., 1997.