

# 광고효과 제고를 위한 동적 Interface 방법에 관한 연구

김경돈\*, 전진호\*\*, 이계성\*\*  
\* 세진T&M 전임 연구원  
\*\* 단국대학교 전자계산학과  
e-mail : jhijy@cs.dankook.ac.kr

## A Study on the Dynamic Interface Method to Increase Advertisement Effectiveness

Kyung-Don Kim\*, Jin-Ho Jeon\*\*, Gye-Sung Lee\*\*  
\* Dept. Research, SeJin T&M  
\*\* Dept of Computer Science, Dan-Kook University

### 요약

사이버 공간에서 활발히 이뤄지고 있는 전자 상거래에 있어서 불특정 다수에게 고정된 광고를 뿌려주는 방식의 광고는 그 효과에 있어 제한이 있다. 본 논문에서는 베이지안 학습법에 기초한 회원 고객의 특성에 따른 분류화를 통한 고객에 따라 타겟광고가 가능한 기법에 대해 연구하고 이를 가능하게 하는 시스템을 제안한다.

### 1. 서론

전자상거래의 시장규모가 날로 그 비중이 증가함에 따라 인터넷 쇼핑물들의 경쟁이 심화되고 있는 실정이다. 현재 사이버몰의 구현과 운영에 대해서는 많은 연구와 방향이 제시되고 있다. 반면 그 공간을 이용하는 주체가 되는 소비자에 대한 연구는 너무나도 미흡하다.

그러므로 여기서는 인터넷의 특징인 정보의 제공자와 수혜자가 상호 작용이 가능하다는 특징에 기초하여 개개인의 성향에 따라 맞춤형 정보를 제공할 수 있다는 점에 착안하여 고객에 따라 타겟 광고가 가능할 수 있도록 하는 기법에 대해 연구하고자 한다. 따라서 본 연구의 목적은 다음과 같다.

첫째, 사용자의 성별, 나이 등에 따라 구매성향이 다름에 따라 이 사용자들을 분류화 할 수 있는 방법을 찾는다.

둘째, 분류화된 사용자들이 관심을 갖는 제품은 어떠한 것인지를 알고자 한다.

셋째, 분류화된 사용자들에 대한 소비성향 내지는 구매의욕이 파악되면 각각의 사용자 부류에 따라 타겟광고가 이루어 질 수 있도록 하는 방안에 대하여 연구하고자 한다.

### 2. 고객분류를 위한 Bayesian model 연구

Bayesian model을 이용한 분류 방법은 감독 분류를 수행하는 방법으로서 교육단계에서는 레이블을 갖는 클래스의 각각의 속성들에 대해 조건부 확률 값을 계산함으로써 이루어지며, 새로운 데이터를 분류하는 과정은 교육단계에서 계산된 조건부 확률 값을 이용해 사후확률을 계산함으로써 진행된다.

Bayesian분류기는 클래스들의 집합  $c_j = C$ 와 속성들의 집합  $a_i = A$ 로 정의된다.

케이스에 대한 레이블이 주어져 있고 속성 값들이 있는 데이터베이스 D가 주어졌을 때, 훈련은 D로부터 주어진 클래스에 대한 속성 값들의 조건부 확률을 계산함으로써 이루어진다. 클래스  $c_j$ 에 대해서 속성  $a_{ik}$ 가 나타날 확률값  $p(a_{ik}|c_j)$ 은 다음 식(1)으로 계산 가능하다.

$$p(a_{ik}|c_j) = \frac{n(a_{ik}|c_j)}{\sum_k n(a_{ik}|c_j)} \quad (1)$$

위의 공식을 이용하여 조건부 확률을 계산함으로써 훈련을 마친 데이터베이스에 대해 두 번째 단계인 새로운 데이터들을 조건에 맞는 클래스로 분류하는 과정은 다음과 같다. 새로운 케이스에 대한 속성 값들의 집합 ( $e_k = [A_1 = a_{1k}, \dots, A_m = a_{mk}]$ )가 주어졌을 때

사후확률은 다음 식(2)과 같다.

$$p(c_j|e_k) = \frac{\prod_{i=1}^n p(a_{ik}|c_j)p(c_j)}{\sum_{h=1}^n \prod_{i=1}^n p(a_{ik}|c_h)p(c_h)} \quad (2)$$

Bayesian 분류기에서는 주어진 클래스들에 대한 속성 값들이 상호 독립적이라는 가정에 의해 다음과 같은 식(3)이 성립한다.

$$p(B|S_1, S_2) = \frac{p(S_1, S_2|B) \cdot p(B)}{p(S_1, S_2)} \quad (3)$$

그러므로, 위의 사후 확률값  $p(c_j|a_{1k})$ 를 구하려면 우선 다음 식(4)를

$$p(c_j|a_{1k}) = \frac{p(a_{1k}|c_j)p(c_j)}{\sum_{h=1}^n p(a_{1k}|c_h)p(c_h)} \quad (4)$$

계산한 후 계산 값을 이용해 그 다음 단계인 식(5)의  $p(c_j|a_{1k}, a_{2k})$  값을 계산한다.

$$p(c_j|a_{1k}, a_{2k}) = \frac{p(a_{2k}|c_j)p(c_j|a_{1k})}{\sum_{h=1}^n p(a_{2k}|c_h)p(c_h|a_{1k})} \quad (5)$$

이러한 과정을 반복함으로써 사후 확률값  $p(c_j|e_k)$ 를 계산할 수 있다. 이와 같이 모든 클래스들에 대해 사후 확률값을 계산한 후 가장 큰 사후 확률값을 가지는 클래스로 주어진 데이터를 분류한다.

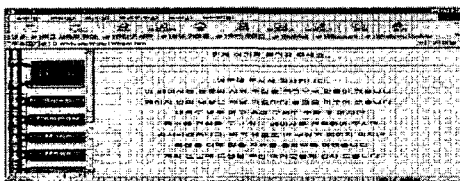
### 3. 고객자료의 수집과 분류 결과

#### 3.1. 사이트 접속을 통한 자료의 수집

본 연구에서의 자료 수집은 내용이 스포츠, 컴퓨터, 음악, 영화, 여행, 건강, 종교에 관한 자료를 수록한 사이트를 통해 고객정보(사용자의 성별, 나이, 성격 등)를 입력받고, 사이트 접속을 통한 로그파일의 수집으로 병행하였으며[1][4], 대상으로는 사이트 접속의 용이함을 위해 19세에서 28세까지의 A대 학생과 B대 학생 중 전산실에서 수업을 하는 학생들을 상대로 조사를 실시하였다.

이를 토대로 나이, 성별, 관심분야에 따른 사용자 부류를 분류하고자 하였는데 성격은 활동적, 차분한, 적극적, 소극적 중 2개 이내로 선택하게 하여 이를 통해 나눠져진 8개의 성격으로 class를 분류하도록 하였다.

다음의 <그림1>은 고객 정보와 로그파일을 얻기 위한 사이트의 메인화면이다.



<그림 1> 자료 수집을 위한 메인 사이트

다음 <그림 2>는 사용자가 접속한 내용을 로그파일을 통해서 자료를 얻은 후, 접속IP별로 정렬을 한 뒤 다시 시간별로 정렬을 하여 개개인의 접속자가 구분이 될 수 있도록 한 것이다.

<그림 2 로그파일의 정렬>

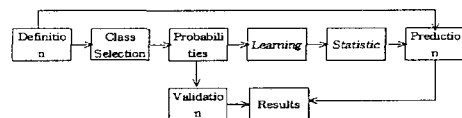
본 연구에서는 <그림 3>에서처럼 가장 먼저 본 페이지에 대해서 가장 관심을 두는 분야라는 가정 하에 '7'에서 '1'까지 가중치를 주었다. 또한 회원정보에서의 성별, 나이와 함께 RoC[1]을 이용해서 고객분류 분석을 수행하기 위한 각각의 케이스에 대한 속성 값으로 사용하였고 회원정보를 통해 얻은 성격에 대한 정보를 레이블로 두어 각각의 케이스를 구성하였다.

<그림 3 로그파일의 가중치 부여>

#### 3.2 데이터 분류

RoC를 이용한 분류는 일부자료를 통해 학습을 한 뒤 그 학습 내용을 가지고 새로운 데이터에 대해서 예측을 하는 것이다. 본 연구에서는 총 184개의 case로 구성되어 있으며 이 중 90%를 학습을 하고 무작위로 추출된 10%에 대해서 분류를 한다. 이러한 과정을 20회 가량 반복적으로 수행해 평균 예측율이 얼마나 되는지를 알아보았다.

RoC의 수행과정은 다음의 <그림 4>와 같다.



<그림 4 RoC의 수행과정>

본 논문에서는 이와 같은 과정을 20회 실시하여 <

1) RoC(Robust Bayesian Classifier)는 베이지안 이론을 근거로 일부자료를 통해 학습을 한 뒤 그 학습 내용을 가지고 새로운 데이터에 대해서 예측을 하는 딱지로서 고객성향 분류에 적합한 틀이다.

표 1>과 같은 결과를 가져왔다.

<표 1> Classifying 예측 결과 (단위: 퍼센트(%))

1회	2회	3회	4회	5회	6회	7회	8회	9회	10회	평균
38.88	50.00	33.33	55.55	38.88	44.44	50.00	27.77	38.88	38.88	
11회	12회	13회	14회	15회	16회	17회	18회	19회	20회	
44.44	33.33	50.00	55.55	55.55	38.88	50.00	38.88	27.77	44.44	42.77

결과적으로 42.77%의 정확한 예측율을 가지고 있는 것을 알 수 있었다. 이것은 클래스를 8개로 나누었을 때 임의의 확률인 1/8의 확률보다 3.42배 많은 것이며, 광고효과를 340% 높은 결과를 가져온다고 할 수 있다.

### 3.3. 고객과 광고의 연관성 분석 및 결과

본 절에서는 사이트의 접속을 통해서 얻어진 자료를 토대로 각 class로 분류된 사용자들의 상품에 대한 지지도를 조사하여 독립성 검정을 통해서 각 class마다 특정 상품에 대해 선호하고 있는지를 파악하였다. 광고에 대한 독립성 검정에서는 범주형 변수를 8개의 각 class와 각 class마다 선택한 11개의 제품군으로 하였고, 8×11의 제품 지지도 분할표는 <표 2>와 같다.

<표 2> 제품 지지도 분할표

	선택된	영양/유형/혼신	패션의류	패션잡화	코스메틱	스포츠/레저	티셔	게임	음반	도서	여행	계
활동적	6	4	4	0	0	4	3	0	1	0	0	22
차분한	6	3	1	2	0	3	4	2	7	1	1	30
적극적	2	1	0	0	0	0	0	0	1	0	0	4
소극적	3	0	2	0	1	0	2	1	0	0	1	10
활동적이고 적극적	6	2	1	0	1	8	4	2	2	2	0	28
활동적이고 소극적	3	3	0	0	0	1	0	0	0	0	0	7
차분하고 적극적	3	0	0	0	0	0	2	0	2	0	0	7
차분하고 소극적	3	0	0	0	0	0	1	3	3	0	3	13
계	32	13	8	2	2	16	16	8	16	3	5	121

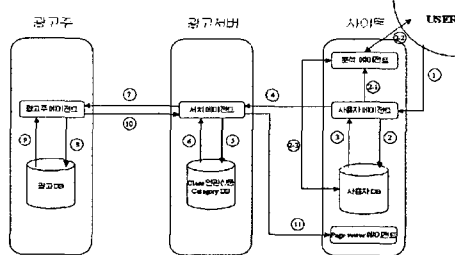
<표 2>에서 상대적으로 적은 선택율을 보여준 패션잡화와 코스메틱, 도서상품권을 소거하여 실험한 결과 5%의 유의 수준에서 각 클래스와 클래스가 선호하는 상품 군과는 서로 연관성이 있음이 증명되었으며, 모든 클래스가 컴퓨터 제품군에 대한 지지도가 높은 것을 알 수 있는데 이것은 표본으로 삼은 대상들이 대학생이라는 특수한 상황 때문이라고 사료된다. 컴퓨터 제품군에 대한 지지도를 제외하면 각 클래스마다 제품군에 대한 지지도가 특징적이라는 것을 알 수가 있다.

지금까지의 실험을 통해서 사용자의 나이, 성별, 관심분야에 따른 데이터를 가지고 사용자들의 성격적인 성향을 예측할 수 있으며, 또한 성격적인 성향에

따라 분류된 클래스들과 각 클래스들이 선호하는 제품군 사이에는 연관성이 있음을 밝혔다.

### 4. 동적 Interface의 설계

위의 실험 결과에 따라 전자상거래 상에서 사용자마다 그들이 선호할만한 제품군에 대해서 성공적인 타겟 광고를 할 수 있음을 말한다. 본 장에서는 실험 결과를 기초로 인터넷상에서 광고효과를 높이기 위한 방법으로 다음 <그림5>과 같은 시스템을 구상하였다.



<그림 5> 동적 광고 Process

<그림 5>에서 사용자 DB에는 사용자에 관한 사용자 프로파일외에 사용자 Class 정보를 담고 있고, 광고서버의 Class 연관상품 category DB에는 각 Class들이 선호할만한 광고에 대한 category 정보와 해당 광고주에 대한 정보를 가지고 있다[2]. Process는 다음과 같이 이루어진다.

우선 사용자가 사이트에 접속하게 되면 사이트상의 사용자 에이전트는 사용자 DB로부터 Class 정보를 추출하고 이를 광고서버의 검색 에이전트에게 맡긴다. 이때 사용자가 기존의 사용자가 아니고 Class가 정의 되어있지 않은 새 사용자라면 사용자 에이전트는 분석 에이전트에게 이를 알린다. 분석에이전트는 새 사용자가 방문하는 page들을 모니터링하고 사용자 DB의 데이터를 바탕으로 사용자를 Class화하고 이를 사용자 DB에 저장한다. 그럼 다시 사용자 에이전트는 사용자 DB로부터 이 새 사용자에 대한 Class 정보를 추출하여 검색 에이전트에게 보내고 검색 에이전트는 사용자 에이전트에게서 받은 클래스 정보에 해당하는 상품광고 Catagory를 추출하여 광고주의 광고주 에이전트에게 각각의 상품정보에 대한 광고파일을 요청한다. 광고주의 광고주 에이전트는 그에 해당하는 광고 파일에 대한 링크정보를 검색 에이전트에게 보내면 검색 에이전트는 이를 다시 사이트의 Page viewer 에이전트에게 보내고 Page Viewer 에이전트는 광고 게재란에 상품광고에 대한 링크를 연결하게 하여 동적 광고를 실현하게 되며 더 간단한 형태의 광고 에이전트를 구성할 수도 있다[3].

## 5. 결론

본 논문에서는 사용자마다 특성을 찾아 Class화할 수 있는 방법에 대해서 연구하고 각 Class마다 구매력을 갖는 제품에는 어떠한 것이 있는지를 파악하고자 하였으며 실험을 통하여 사용자의 나이, 성별, 관심분야에 따라 사용자를 성격적인 성향으로 나누고 이를 토대로 새로운 사용자에게 대해서도 성공적인 분류를 수행할 수 있음을 밝혀 사용자들을 성격적이 성향에 따라 Class화하는 것은 의미가 있음을 알아냈고 그렇게 분류된 사용자 Class들은 나름대로의 제품에 대한 구매성향이 있음을 파악하였다. 하지만 본 논문에서는 표본으로 한 대상이 대학생으로 한정적이었기 때문에 폭넓은 자료에 대한 분석은 실시할 수 없었다.

향후 좀더 폭넓은 대상을 표본으로 한 연구가 이루어져야 할 것이다. 또한 온라인 학습을 통해 사용자 부류를 나눌 수 있는 방안에 대한 연구가 좀더 이루어진다면 전자상거래 분야에 있어서 개개인의 고객에게 차별화 된 맞춤 광고를 제공하는데 도움이 될 것이다.

## [참고문헌]

- [1] 이화영, "표준 로그파일을 이용한 웹마이닝에 관한 연구", 한국과학기술원, 2000.
- [2] 이경전, "전자상거래 소프트웨어 에이전트", 정보처리학회지 제6권 제1호, 1999.
- [3] Kim, J., Kim, Y., Yoo, K., Kang, T., Han, S., Jun, H., "Personalized Advertisement on Electronic Commerce System Using Unsupervised Learning", Joint Conference of KORMS/KIES, Spring 1998.
- [4] Cooley, R, Mobasher, B. and Srivastava, J. (1999), "Data preparation for mining world wide web browsing pattern", journal of knowledge and Information Systems, Vol. 1, No.1.