

# 무선인터넷 환경에서의 개인화상품추천 에이전트

이승화<sup>o</sup> 이은석

성균관대학교 정보통신공학부

jbmania@selab.skku.ac.kr<sup>o</sup>, eslee@ece.skku.ac.kr

## A Personalized Product Recommendation Agent on Mobile Internet

SeungHwa Lee<sup>o</sup> Eunsoek Lee

School of Information and Communication Engineering, SungKyunKwan University

### 요약

본 논문에서는 무선인터넷 환경에 적합한 개인화된 상품추천에이전트를 제안한다. 기존에 유선인터넷상의 많은 개인화 추천시스템에서는 초기 사용자 모델링을 위해 사용자에게 수많은 질문을 하고 응답을 요구하였다. 그러나 이러한 방식은 무선인터넷 환경에서 정보 전송량에 따른 높은 사용요금을 고려할 때 적용하기 힘든 방식이다. 본 제안 시스템은 사용자의 Social data를 이용하여 사용자를 비슷한 연령과 성별 그룹으로 나누고, 해당 그룹에서 구매율이 높은 상품을 우선 제시한 후, 사용자 행동을 모니터링 하여 암시적(Implicit)피드백을 통해 프로파일을 생성함으로써, 번거로운 질의-응답 과정 없이도 초기 사용자 모델링을 수행할 수 있다. 프로파일 생성 이후에는 이를 기반으로 하여 사용자 유사한 취향을 가진 그룹으로 다시 군집화한 후 협력적 추천을 하게 되며, 프로파일에는 해당 상품의 최종 카테고리명과 키워드를 수집함으로써, 상품의 브랜드와 규격정보를 반영한 추천이 가능하다. 또한 추천 상품과 사용자의 구매데이터와의 비교를 수행하여 사용자가 해당상품을 구매하였을 경우, 상품에 대한 취향정보는 그대로 유지하고 관련 상품을 추천하되, 구매한 상품이 중복 추천되지 않도록 하였다. 시스템 평가를 위해 프로토타입을 구현하여, 다수의 사용자에게 시스템을 이용하며 관심품목을 체크하도록 하였고, 추천횟수가 반복되며 히트율이 증가하는 결과를 통해 시스템의 학습속도와 성능을 평가하였다. 그리고 쇼핑몰에서 구매경험이 있는 사용자의 기존 구매데이터와 Social data를 이용한 초기 제시상품을 역으로 비교하여 오랜 시간과 비용 발생 없이도 초기 프로파일 생성의 유효성을 증명하였다.

## 1. 서론

정보통신기술과 인터넷의 급격한 발전을 통해 우리는 수많은 정보를 보다 빠르고 손쉽게 접할 수 있게 되었으며, 이러한 인터넷의 확산과 함께 전자상거래도 폭발적으로 증가하고 있다. 또한 최근에는 이동통신기술의 발전을 통해서 셀룰러폰이나 PDA와 같은 Handheld 디바이스의 보급이 확산되었고, 그로 인해 기존의 유선상의 인터넷 서비스는 무선 환경으로 점차 확대되고 있다. 그러나 이러한 변화에도 불구하고, 아직 무선 환경의 전자상거래 서비스 이용률은 그다지 높지 않은데, 그 이유로는 작은 화면과 적은 메모리 용량과 같은 Handheld 디바이스의 제한된 성능으로 인해 제공받는 정보가 한정되어 있다는 점을 들 수 있으며, 작은 키패드를 통한 조작의 번거로움과 사용량에 따라 부과되는 높은 이용요금도 무선 전자상거래의 확산을 막는 문제점으로 지적되고 있다.

본 논문에서는 이러한 문제점을 해결하기 위해 사용자의 취향을 자동으로 파악하여 사용자에게 적합한 상품을 추천해주고, 전송되는 정보의 양을 효율적으로 줄이는 무선인터넷 환경에 적합한 개인화된 상품추천 에이전트를 제안한다.

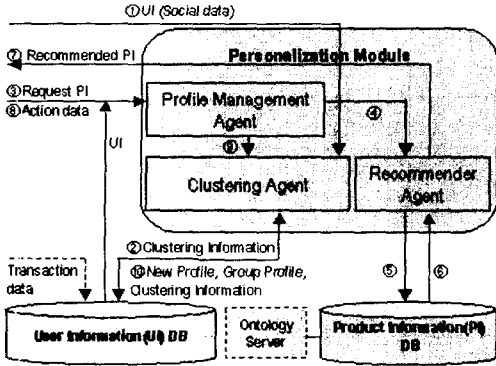
본 논문의 구성은 다음과 같다. 2장에서는 관련연구로 기존의 유선인터넷상의 개인화 시스템 연구를 살펴보고, 3장에서는 제안시스템에 대해 세부적으로 설명하였다. 4장에서는 실험을 통한 평가를 하였으며, 결론 및 향후과제를 5장에 기술하였다.

## 2. 관련연구

기존 유선 인터넷상에서 개인화된 상품추천을 위한 많은 연구가 있었다. 개인화 서비스를 제공하는 대표적인 온라인 상점으로 Amazon.com[2,3]을 들 수 있으며, 추천과정은 사용자모델링(초기 프로파일 생성) → 추천, 필터링 → 학습(프로파일갱신)의 순서로 진행된다. 이 중, 사용자 모델링 과정에서는 사용자에게 다수의 질문을 하고, 명시적인(Explicit) 피드백을 통해 초기 프로파일을 생성한다. 이러한 방법은 정확성과 학습속도가 빠른 장점이 있지만, 사용자를 번거롭게 할 수 있고, 무선인터넷 환경에서는 정보 전송량에 따른 높은 사용요금 등을 고려할 때 적용하기 힘든 방식이다. 따라서 본 논문에서는 사용자의 Social data를 통해 비슷한 연령과 성별을 가진 사용자 그룹에서 구매율이 높은 상품을 먼저 제시한 후 사용자의 행동을 모니터링 함으로써, '클릭'과 같은 암시적(Implicit) 피드백을 통해 초기 프로파일을 학습하였다.

## 3. 제안시스템

본 논문에서 제안하는 개인화 시스템의 구성과 작동 순서는 (그림 1)과 같다. 먼저 사용자를 주민번호 분석을 통하여 <표 1>과 같이 비슷한 연령과 성별로 그룹화 한다. 그 후, 사용자의 첫 번째 추천 상품 요청이 있을 때, 해당 그룹의 구매데이터를 분석하여 가장 구매율이 높은 상품을 위주로 추천해준다.



(그림 1) 개인화 모듈의 구성과 작동 순서

| No | 이름  | 전화번호       | 주민번호           | 분류 |
|----|-----|------------|----------------|----|
| 1  | 홍길동 | 0101112222 | 720808-1XXXXXX | E  |
| 2  | 박세리 | 0102223333 | 730808-2XXXXXX | F  |
| 3  | 박찬호 | 0103334444 | 700808-1XXXXXX | E  |
| 4  | 김미현 | 0104445555 | 830808-2XXXXXX | D  |
| 5  | 선동열 | 0105556666 | 860808-1XXXXXX | A  |
| :  | :   | :          | :              | :  |

A그룹:10대 남성사용자 B그룹:10대 여성사용자  
 C그룹:20대 남성사용자 D그룹:20대 여성사용자  
 ... K그룹:60대 이상사용자

<표 1> 나이와 성별을 이용한 그룹화

첫 번째 추천이 이루어진 후에는, 사용자의 행동을 모니터링 하여 암시적인 피드백을 얻고, 이를 통해 초기 프로파일을 생성한다. 이때 긍정적인 피드백은 상품정보의 '클릭'으로 판별하며, 무반응을 보인 상품은 부정적 응답으로 판별한다. 이러한 프로파일의 생성과 학습에는 Rocchio의 적합성 피드백 알고리즘[4]을 응용하여 사용하며, 이를 식으로 나타내면 다음과 같다.

$$P' = P + \frac{\alpha}{N} \sum RP^+ - \frac{\beta}{M} \sum RP^- \quad (\text{식 1})$$

여기에서  $P'$ 는 갱신된 프로파일,  $P$ 는 기존 프로파일,  $RP^+$ (Recommended Product)는 긍정적 피드백을 보인 상품,  $RP^-$ 는 부정적 피드백의 상품,  $N, M$ 은  $RP^+$ 와  $RP^-$ 의 개수를 나타내며,  $\alpha$ 와  $\beta$ 는 각 항의 가중치로 이 값을 조정함으로써 새로운 피드백에 대하여 학습하는 성향을 조정할 수 있다.

위와 같은 과정을 통해 초기 프로파일이 생성된 이후에는, K-Means 알고리즘[5]을 이용하여 사용자들을 보다 유사한 취향을 가진 그룹으로 다시 군집화한다. K-Means 알고리즘은 거리 기반 군집화 방법으로, 사용자의 프로파일을 다차원 공간상의 점으로 표시하고, 거리를 계산함으로써 전체 사용자들의 집합을  $k$ 개의 군집으로 나눈다. 사용자  $a$ 와  $k$  사이의 거리는 (식 2)과 같이 계산하며,  $a_i$ 는 사용자  $a$ 의 속성  $i$ 에 대한 선호도 값을 의미한다.

$$d_{a,k} = \sqrt{\sum_i (a_i - k_i)^2} \quad (\text{식 2})$$

군집화 이후에는 그룹 내 사용자의 Transaction data를 이용한 협력적 추천이 이루어지게 된다.

프로파일은 벡터공간표현방법(Vector Space representation)으로 표현되며, 상품에 대한 다양한 정보를 담기 위해, 상품명의 키워드와 함께 최종 카테고리명이 저장된다. (그림 2)에 그 예가 나와 있으며, 이를 통해 사용자가 선호하는 상품과 함께 브랜드나 규격과 같은 속성도 파악할 수 있게 된다. 그리고 프로파일과 PI DB의 최종 카테고리명의 비교를 통해, 같은 카테고리명이 존재하는 경우 해당 카테고리에서 가장 구매율이 높은 상품을 추천 상품에 포함하게 된다. 따라서 사용자가 상품을 구매하였을 경우에도 해당 상품의 취향 정보는 그대로 가지고 카테고리내의 관련 상품이나 업그레이드 상품을 계속 추천해줄 수 있으며, 구매데이터와의 비교를 수행하여, 구매한 상품은 중복 추천되지 않도록 하였다.

\*추천상품검색\*

- 1.HP프린터 데스크젯
- 2.아이리버 256MB <click>
- 3.쿠쿠 알락밥

root/electronics/av/mp3/pid201205 => mp3

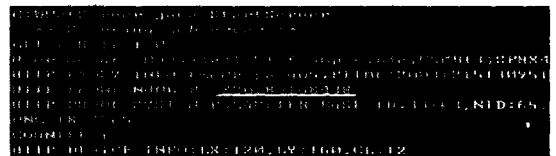
Profile = {(아이리버, 0.5), (256MB, 0.5), (MP3, 0.5)...}

(그림 2) 프로파일에 추가되는 상품명 키워드와 카테고리명

위의 두 가지 방식을 혼용하여 중복되지 않도록 10개의 상품을 추천하게 된다.

#### 4. 구현 및 평가

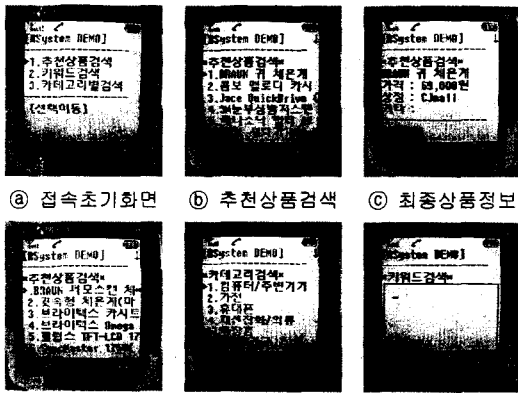
본 시스템은 윈도우 기반에서 자바를 이용하여 자체적으로 웹 서버를 구현하였으며, 각 에이전트들은 서블릿과 JSP로 제작되었다. DB는 MS-SQL 2000을 이용하였으며, 단말기로 전송되는 결과화면은 mHTML을 이용하였다. 구현된 웹 서버는 사용자가 접속하였을 때, 서버 쪽으로 전송되는 단말기의 헤더정보를 이용하여 개인 식별을 하는 기능을 갖고 있으며, 이를 통해 사용자가 작은 키패드를 이용하여 일일이 사용자 ID와 비밀번호를 입력하던 번거로운 절차를 간소화하였다.



(그림 3) 서버로 전송되는 단말기의 헤더정보

본 시스템은 추천상품검색, 키워드검색, 카테고리검색의 세 가지 형식의 검색을 지원하며, (그림 4.㉔)은 이러한 검색방식을 선택할 수 있는 접속초기화면이다. 추천 상품검색을 선택하여, 초기 추천 상품이 출력된 화면이 (그림 4.㉕)에 나타나 있으며, 상품정보의 클릭으로 최종 상품 정보까지 검색한 화면이 (그림 4.㉖)이다. 이때 클릭

한 상품의 키워드와 최종카테고리가 추출되어 사용자의 프로파일에 반영되며, 여기에서는 유아-아동용품으로 분류된 상품이 클릭 되어, 'BRAUN', '귀', '체온계'과 같은 키워드와 함께 상품 경로 root/living/baby/pid600102 중 'baby'라는 카테고리명이 프로파일에 저장된다. (그림4,㉔)에는 이러한 내용이 반영되어 관련 상품이 추가로 추천되는 것을 보여주고 있다. (그림 4,㉕,㉖)는 각각 키워드형 검색과 카테고리 형 검색화면을 보여주고 있다.



㉔ 접속초기화면 ㉕ 추천상품검색 ㉖ 최종상품정보  
 ㉗ 개인화된 추천 ㉘ 카테고리형 검색 ㉙ 키워드형 검색  
 (그림 4) 시스템의 구현화면

본 시스템의 평가는 쇼핑몰을 이용한 경험이 있는 사용자 100여명과 그들의 구매데이터, 그리고 수집된 500여개의 상품데이터를 이용하였다. 그리고 초기에 제시되는 각 그룹별 인기상품목록은 국내 쇼핑몰인 Cjmall에서 제공하는 데이터를 이용하였다. 실험은 초기 프로파일 생성 시에 제시하는 Social data에 기반 한 상품 추천의 유효성과 추천시스템의 전체적인 성능평가의 두 방향으로 진행되었다.

우선 초기 상품 추천의 유효성은 실험에 참가한 사용자들의 구매데이터와 제시되는 상품의 비교를 통해 이루어졌으며, 평가 방법으로는 정보검색분야에서 보편적으로 사용되는 평가척도인 Precision을 응용하여 사용하였다. Precision은 전체 추천된 상품 개수 중에서 사용자가 선호하는 상품 개수의 비율로, 여기에서는 구매데이터에 기록된 상품을 사용자가 선호하는 상품으로 간주 하였다.

$$\text{Precision} = \frac{\text{추천된 상품과 일치하는 구매데이터 상품 수}}{\text{Social data에 기반한 초기 추천 상품}} \quad (\text{식 } 3)$$

<표 3> Precision 측정결과

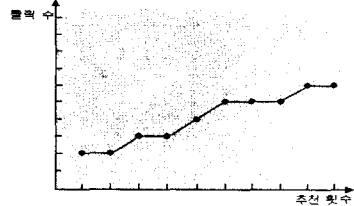
| 추천 A | 추천 B | 추천 C |
|------|------|------|
| 0.22 | 0.18 | 0.08 |

- 추천 A : Social data 기반 상품추천
- 추천 B : 전체에서 구매율 높은 상품추천
- 추천 C : 무작위 상품 추천

실험결과, 두 데이터 간에 일치하는 상품은 평균 2건 이상이었으며, 무작위로 상품을 제시하는 것보다 훨씬

높은 결과를 보였다. 또한 기존에 발생한 데이터를 이용한 본 실험을 통해, 새로운 구매데이터가 쌓일 때까지의 오랜 시간과 비용발생 없이도 성능을 평가할 수 있었다.

두 번째 실험은 유선 웹 페이지를 통해 동일한 추천시스템을 만든 후, 100명의 사용자에게 시스템을 반복 이용하여 관심품목을 체크하도록 하였다. 그 결과, 추천횟수가 반복되면서 히트율이 증가하는 것으로 시스템의 학습 성능을 평가하였다.



(그림 5) 추천횟수 반복에 따른 클릭 수 증가

### 5. 결론 및 향후과제

우선 인터넷의 특성상 단일기의 제한된 성능을 고려한 사용자의 편의성 향상과 개인화된 서비스의 제공은 매우 절실히 요구되는 사항이다. 본 논문에서는 개인화된 추천 시스템을 우선 환경에 맞게 제안하여, 사용자에게 보다 유용한 상품을 추천함과 동시에, 불필요하게 전송되는 정보의 양을 효율적으로 줄여 우선 환경에서 전송 패킷 단위로 부과되는 사용요금의 감소를 유도하였다.

이러한 제안 시스템의 기능은 모바일 전자상거래를 활성화 시키는 효과를 가져올 것으로 기대되며, 향후 연구 방향과 연구과제로는 다음과 같은 내용이 필요하다.

- 보다 다양한 추천모델에 대한 연구
- 사용자의 반응에 따라 추천 상품 리스트의 추천비율과 순위를 자동 조정하는 연구
- 사용자가 클릭한 상품의 가격정보와 상품이 저장된 카테고리의 평균가격과의 비교를 통한 사용자의 가격 민감도 정보수집 및 이를 반영한 추천시스템
- 유선환경의 쇼핑몰과의 통합으로, 유선에서 이루어진 상품검색을 통해 사용자의 취향을 파악하고, 핸드폰으로 상품을 추천하여, 최종 구매결정을 유도하는 유·무선이 연계된 서비스와 위치정보를 활용한 오프라인 쇼핑몰과의 연계된 서비스.

### [참고문헌]

- [1] Eunseok Lee, "A Next Generation Intelligent Mobile Commerce System", LECTURE NOTES IN COMPUTER SCIENCE Springer-Verlag, pp.131-141, Feb. 2004
- [2] Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations : Item-to-Item Collaborative Filtering", IEEE Internet Computing, Vo.7, No.1, pp. 7680, Jan.2003
- [3] <http://www.amazon.com>
- [4] Young-Woo Seo, Byoung-Tak Zhang, "A Reinforcement Learning Agent for Personalized Information Filtering", In Proceedings of International Conference on Intelligent User Interface '2000(IUI-2000), pp.248-251, New Orleans, Louisiana, USA, Jan.9-12, 2000.
- [5] Lyle H. Ungar and Dean P. Foster, "Clustering Methods for Collaborative Filtering", Proceeding of the 1998 Workshop on Recommendation Systems, pp.114-129. 1998.