

인터넷 쇼핑몰을 위한 하이브리드 상품 추천 시스템

천인국^o

순천향대학교 정보기술공학과

Hybrid Product Recommender System for Internet Shopping Mall¹

In-Gook Chun^o

Division of Information Technology Engineering, SoonChunHyang University

chunik@sch.ac.kr,

요 약

본 논문은 인터넷 쇼핑몰에서의 효율적인 상품 추천 시스템의 구조를 제안한다. 본 상품 추천 시스템은 상품 추천의 2 가지 방법인 지식기반 상품 추천 방법과 collaborative filtering 과 을 혼합하였으며 먼저 고객에게 질문을 던져서 고객의 요구 조건을 수집한 다음, 요구 조건과 상품 데이터베이스에 저장된 상품 정보와 일치도를 계산하여 추천 후보 상품 리스트를 생성한다. 이 추천 상품 리스트에 속하는 상품에 대해서는 다시 collaborative filtering 방법이 적용된다. 즉 비슷한 취향을 가지는 고객들이 높이 평가하는 제품들을 최종적으로 고객들에게 추천하게 된다. 이 방법은 기존의 방법들이 모두 특정한 상품 카테고리에 대해서만 효과적이던데 비하여 제안된 방법은 모든 상품 카테고리에 적용할 수 있으며 collaborative filtering 방법을 후보 추천 상품에 대해서만 적용시킴으로써 이 방법의 단점인 많은 계산량을 줄일 수 있다. 제안된 시스템은 EJB(Enterprise Java Beans)를 사용하여 컴포넌트로 구현되었으며 이동통신기기 카테고리에 대하여 시험 구현되었다.

1. 서론

상품 추천 시스템은 인터넷 쇼핑몰에서 방문객에게 적절한 상품을 추천해주는 역할을 한다. 가장 널리 쓰이는 기법은 collaborative filtering 이다[6][3]. collaborate filtering 은 고객의 상품에 대한 평가 결과, 과거 구매 기록이나, 웹페이지 방문 정보 등을 분석하여 비슷한 취향을 가진 다른 고객들이 많이 선택한 상품을 추천해주는 시스템이다. collaborative filtering 은 인간의 취향이 서로 연관이 있다는 가정에서 출발한다. 따라서 어떤 고객과 비슷한 경향의 사람들이 좋아하는 물건이면 그 고객이 구입할 가능성도 높다고 보는 것이다. collaborative filtering 은 주로 책이나 음반, 영화들을 선택하는데 유용한 방법으로 알려지고 있다. 또 하나의 추천 기법은 지식기반의 추천 시스템이다[2][8]. 지식기반의 추천 시스템은 고객에게 원하는 상품에 대한 요구사항을 질문하고 그 답변을 바탕으로 하여 고객의 요구 조건에 맞는 상품을 추천하여 추천한다. 이를 위해서는 상품 영역의 지식이 추론이 가능하게끔 지식

이므로 정리되어 있어야 한다. 지식 기반 추천이 적절한 상품 분야로는 집, 자동차, 컴퓨터 등이 있다. 이 분야의 사이트로는 <http://www.personallogic.com>에 있는 PersonalLogic 사이트가 대표적이다. 각 방법은 각각의 장점과 단점을 가지고 있다. collaborative filtering 은 완전히 자동화된 추천이 가능하고 고객의 취향에 대한 데이터가 수집될수록 더 수준 높은 추천을 할 수 있다. 반면에 단점으로는 추천을 시작하려면 상당히 많은 고객의 취향 데이터베이스가 필요하다는 것이다. 과거의 취향 기록을 사용하기 때문에 고객의 취향이 바뀌면 유용한 추천이 되지 않을 수도 있다. 또한 추천 상품을 결정하는데 많은 계산 시간이 필요하다. 지식기반의 추천 방법은 고객 취향의 초기 데이터베이스를 필요로 하지 않는다. 자동차나 집, 컴퓨터 등의 상품에 대해서는 본질적으로 지식 기반의 방법이 더 적절하다. 예를 들면 자동차를 사고자 하는 고객이 있을 경우에, 자동차의 여러 가지 특징, 예를 들면 차종, 크기, 가격 등의 상품 특징이, “다른 사람이 어떤 자동차를 샀느냐” 보다 더 중요하다. 반면에 지식 기반 방법은

¹ 본 연구는 정보통신부의 ITRC 사업에 의해 수행된 것 임

상품에 대한 지식을 가지고 있어야 하고 이들 지식이 추론이 가능한 상태로 저장되고 조직화되어야 한다. 따라서 지식 공학자를 필요로 한다[8]. 본 논문에서는 이들 2 가지 방법을 동시에 사용하는 상품 추천 시스템을 제안하고자 한다. 제안된 방법에서는 먼저 사용자의 상품에 대한 요구 조건을 질의한다. 고객의 요구 조건이 수집되면 요구 조건에 일치하는 상품을 상품 데이터베이스에서 찾는다. 여기서는 요구조건을 100% 만족하는 상품만 찾는 것이 아니라 요구 조건을 일정 수준 이상 만족하면 추천 후보 리스트에 포함시킨다. 만약 사용자가 질의에 대하여 대답을 하지 않으면 추천 후보 리스트는 해당 카테고리의 전체 상품이 된다. 사용자가 더 많은 질의에 답변을 할수록 추천 범위가 줄어들게 되고 따라서 추천 후보 상품 리스트의 크기가 작아지게 된다. 추천 후보 상품 리스트가 완성되면 다시 이 리스트에 속하는 상품에 대하여 collaborative filtering 알고리즘이 수행된다. 현재의 고객과 비슷한 취향의 다른 고객을 찾은 다음, 이들이 높게 평가한 상품을 최종 추천 리스트에 포함시킨다.

본 논문에서 제안된 시스템은 먼저 지식기반 방법과 collaborative filtering 방법을 혼합한 시스템으로서 양쪽의 장점을 취하고자 했다. 즉 상품의 종류에 상관없이 사용할 수 있는 상품 추천 시스템을 만들고자 했으며, collaborative filtering 방법을 추천 후보 상품에 대해서만 적용시킴으로써 이 방법의 단점인 많은 계산량을 줄이고자 하였다. 제안된 시스템은 객체 지향 컴포넌트 기반의 쇼핑몰에서 JAVA 기반의 EJB(Enterprise Java Beans)를 사용하여 컴포넌트 형태로 구현되었으며 핸드폰 카테고리에 대하여 대하여 시험적으로 구현되었다.

2. 상품 추천 시스템의 구조

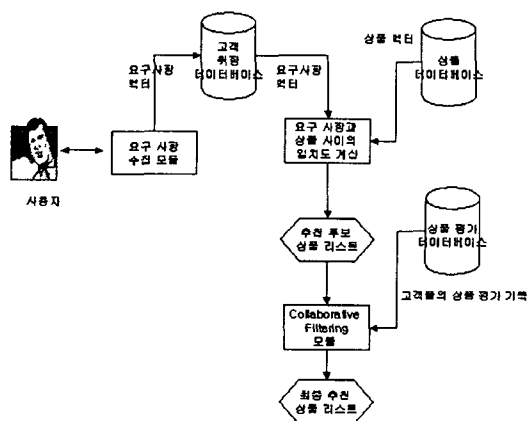


그림 1. 상품 추천 시스템의 전체 구성도

상품 추천 시스템은 3 가지의 모듈로 구성되어 있다.

첫 번째 모듈은 고객으로부터 요구 사항을 질문하고 그 답변을 얻어내는 요구 사항 수집 모듈이다. 요구 사항 수집 모듈에서는 고객의 요구 사항이 출력되고 이것은 고객 취향 데이터베이스에 기록된다. 요구 사항 수집 모듈에서는 상품 데이터베이스를 참조하여 상품의 특징을 웹페이지로 사용자에게 제시하고 각 항목에 대한 사용자의 응답을 기록한다. 모든 항목의 초기값은 “상관없음”이며 사용자가 선택을 하지 않는 항목들은 “상관없음”으로 남아있으며 이 항목들은 요구사항과 상품과의 일치도를 계산할 때 제외된다. 고객은 요구 사항 입력 도중에 언제든지 입력을 멈추고 다음 단계로 진행할 수 있으며 그럴 경우에는 현재까지 고객이 입력한 정보만을 가지고 추천 후보 상품 리스트를 만든다. 따라서 전체 상품이 추천 후보가 될 수도 있다. 그럴 경우에는 기존의 collaborative filtering 과 동일한 성능을 보이게 된다.

두 번째 모듈은 고객의 요구사항과 상품 특징 정보 사이의 일치도를 계산하는 모듈이다. 일치도는 요구 사항과 상품을 모두 벡터로 보고 두 벡터 사이의 근접도를 측정하는 것이다. 사용할 수 있는 측정값으로는 Pearson correlation 과 Euclidian distance, cosine projection 등이 있다. 이들 중 하나를 선택하는 문제는 휴리스틱의 문제이다[7]. 본 논문에서는 cosine projection 을 선택하였다. 세번째 모듈은 취향이 비슷한 고객들이 높게 평가하는 상품을 결정하는 collaborative filtering 모듈이다. 이미 계산된 각 고객 사이의 유사도에 의하여 각 추천 후보 상품에 대하여 추천값을 계산한다. 가장 추천값이 높은 순으로 N 개의 상품을 고객에게 제시하게 된다. 상품 데이터베이스에는 그 상품의 대분류 코드와 소분류 코드, 상품의 특징 등이 포함되어 있다. 상품의 각각의 특징은 데이터베이스에 다음과 같은 구조를 사용하여 입력되어 있다.

이름	타입	값의 범위	값	특징의 분류	설명
resolution	NUMERIC	1-5	3	Performance	해상도
round design	NUMERIC	BINARY	1	Shape	디자인
...					
...					

상품 특징의 타입에는 BINARY, NUMERIC 이 있다. BINARY 타입은 1 또는 0의 값을 가지며 그 특징의 존재 여부를 나타낸다. NUMERIC 타입은 보통 1에서 5까지의 범위를 가지며 그 특징을 얼마나 지원하는지를 숫자로 나타낸다. 1 이면 전혀 지원하지 않는 것이고 5 이면 그 특징을 완벽하게 지원함을 의미한다. 특징의 분류는 특징을 몇 개의 카테고리로 나누는 것이다. 현재는 Size, Weight, Texture, Ingredient, Color, Shape, Packing, Performance, Content, Others 등으로 구분한다.

상품 평가 데이터베이스는 여러 가지 정보에 의하

여 작성될 수 있다. 첫번 고객이 명시적으로 각 상품에 대하여 평가를 할 수 있다. 대개 1에서 5까지의 수치를 이용하며 숫자가 증가할수록 긍정적인 평가이다. 두번 는 각 고객이 과거에 구매한 상품은 이미 고객으로부터 평가가 되었다고 가정하는 것이다. 세번째는 고객의 웹페이지 상에서의 여러가지 행동을 분석하여 해당 상품에 대한 관심도나 호감을 그 상품에 대한 평가점수로 기록할 수도 있다.

3. 상품 추천 알고리즘

상품에 대한 요구 사항을 받아서 이를 만족하는 상품을 찾는 모듈에서는 상품의 특징에 의하여 정의되는 벡터 공간에 고객의 요구 사항과 상품을 표현하였다. 요구 사항과 상품 사이의 일치도는 요구 사항 벡터와 상품 벡터간의 cosine projection으로 나타낸다. 따라서 고객 m 과 상품 n 과의 일치도 점수 \bar{S}_{mn} 는 다음과 같은 식에 의하여 계산된다.

$$S_{mn} = \frac{\bar{C}_m \cdot \bar{P}_n}{\|\bar{C}_m\| \|\bar{P}_n\|}$$

여기서 \bar{C}_m 은 고객 m 의 요구사항 벡터이고 \bar{P}_n 은 상품 n 의 상품 특징 벡터이다. 여기서 \cdot 연산자는 inner (dot) product 를 의미한다.

Collaborative filtering 모듈에서는 다음과 같은 식을 이용하여 고객 m 의 상품 n 에 대한 추천 점수를 계산한다.

$$C_{mn} = \bar{C}_m + \kappa \sum_{i=1}^k w(m, n) (C_{in} - \bar{C}_i)$$

여기서 C_{mn} 은 고객 m 의 상품 n 에 대한 추정 점수이고 \bar{C}_m 은 고객 m 의 평균 평가 점수, $w(m, n)$ 은 고객 m 과 고객 n 의 거리, 연관도, 유사도 등을 의미한다. k 은 해당 상품 평가에 참여한 고객의 숫자이다. κ 는 $w(m, n)$ 의 합의 값이 1 이 되게 만드는 normalizing factor 이다.

$w(m, n)$ 을 계산하는 방법은 여러 가지가 있다. GroupLens 프로젝트에서 쓰였던 Pearson correlation coefficient, cosine coefficient 등 여러 가지가 있을 수 있다[8]. 여기서는 앞에서 사용되었던 cosine coefficient 를 사용하였다.

4. 상품 추천 시스템의 구현

제안된 상품 추천 시스템은 이동 통신 기기인 휴대폰에 대하여 구현되었다.

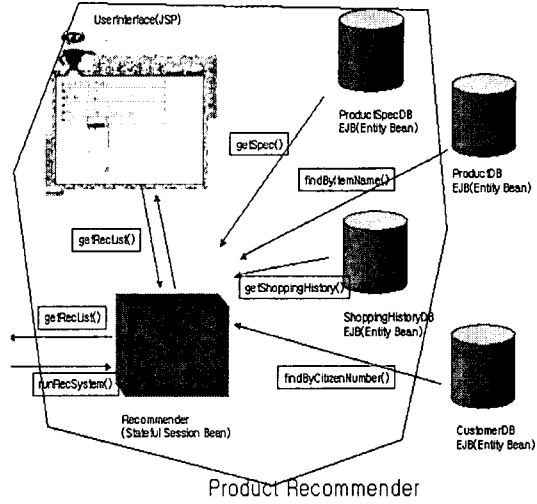


그림 2. 상품 추천 시스템의 흐름도

먼저 휴대폰의 각종 특징을 데이터베이스로 정리하였으며 이 데이터베이스를 바탕으로 고객의 요구사항을 질문하는 웹페이지를 JSP(Java Server Script)를 이용하여 제작하였다. JSP 는 Session EJB(Enterprise Java Bean)로 구현된 상품 추천 시스템 컴포넌트를 호출하게 되며 여기서 각종 데이터베이스를 Entity EJB 를 통하여 접근하여 최종적으로 상품을 추천하게 된다. 다음은 추천 과정에서 나타나는 화면들이다.

■ 색상 [베니시] □ (색상이 없을경우 가장 가까운 색상을 추천해 줍니다)

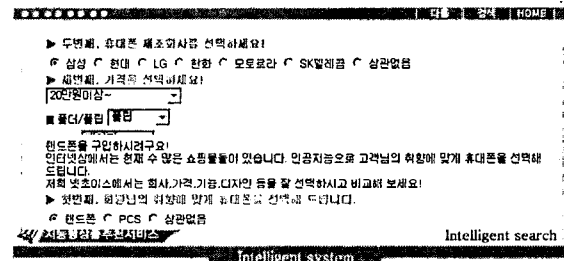


그림 3. 상품 추천 시스템의 처음 화면

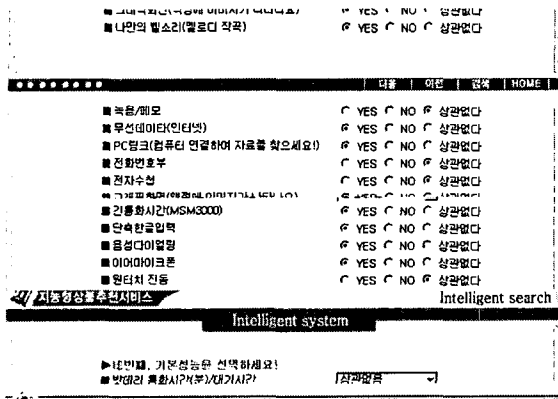


그림 4. 상품 추천의 중간 과정 화면

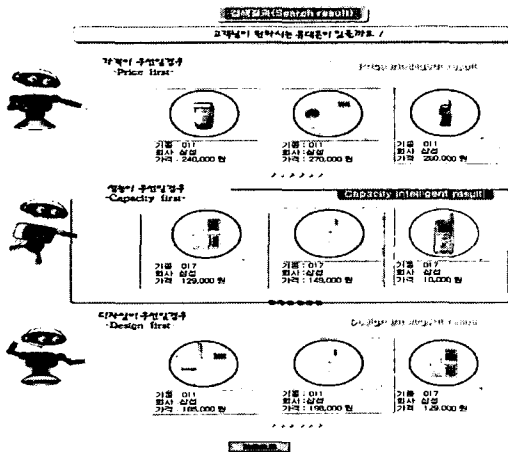


그림 5. 상품 추천의 결과 화면

5. 결론

본 논문에서는 collaborative filtering 의 개념과 지식 기반 추천 시스템을 동시에 지원하는 하이브리드 추천 시스템을 제안하였다. 먼저 지식 기반 추천을 거친 다음 그 결과가 collaborative filtering 모듈로 연결된다. 지식 기반의 추천 모듈에서는 먼저 사용자에게 원하는 상품의 요구 조건을 질문한 다음, 요구 조건에 근접하는 상품을 상품 데이터베이스에서 찾는다. 요구 조건을 어느 정도 만족시키면 그 상품은 추천 후보 리스트에 포함시킨다. 추천 후보 상품 리스트가 완성되면 collaborative filtering 알고리즘이 수행된다. 먼저 현재의 사용자와 비슷한 취향의 다른 사용자들로부터 높게 평가된 상품을 찾은 다음, 이 상품들을 최종 추천 리스트에 포함시킨다. 제안된 시스템은 지식 기반 방법과 collaborative filtering 방법을 혼합한 시스템으로서 양쪽의 장점만을 취하고자 했다. 기존의 추천 시스템들이 한정된 상품에 대해서만 효과적인데 반하여 제안된 추천 시스템은

책이나 음반, 자동차, 컴퓨터 등의 모든 상품 카테고리에 사용할 수 있으며 collaborative filtering 방법을 추천 후보 상품에 대해서만 적용시킴으로써 수행하여야 하는 계산량을 줄였다. 제안된 시스템은 객체 지향 컴포넌트 기반의 쇼핑몰에서 JAVA 기반의 EJB(Enterprise Java Beans)를 사용하여 컴포넌트로 구현되었으며 핸드폰 카테고리에 대하여 대하여 시험 구현되었다.

6. 참고문헌

[1] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, "Analysis of Recommendation Algorithms for E-Commerce", GroupLens Research Group / Army HPC Research Center, Department of Computer Science and Engineering University of Minnesota, 2000.

[2] Burke, R. "Integrating knowledge-based and collaborative-filtering recommender systems". AAAI Workshop on Artificial Intelligence for Electronic Commerce WS-99-01, AAAI Press, Menlo Park, California, 1999.

[3] Jack Breese, David Heckerman and Carl Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, Morgan Kaufmann Publisher, July, 1998.

[4] Mark Claypool, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin (WPI) "Combining Content-Based and Collaborative Filters in an Online Newspaper", ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation, 1999.

[5] Mark O'Conner, Jon Herlocker "Clustering Items for Collaborative Filtering ", ACM SIGIR '99 Workshop on Recommender Systems, August 19, 1999.

[6] Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P. and Riedl, J., Grouplens, "An open architecture for collaborative filtering of netnews". Proceedings of the ACM Conference of Computer Supported Cooperative Work, pp. 175-186, 1994.

[7] R.D. Lawrence, G.S.Almasi, V.Kotlyar, M.S.Viveros, and S.S.Duri, "Personalization of Supermarket Product Recommendations", IBM Research Report, RC 21792, 2000.

[8] Thomas Trans and Robin Cohen, "Hybrid Recommender Systems for Electronic Commerce", University of Waterloo, 1999.

[9] Recommender System, Papers from the 1988 Workshop Technical Report WS-98-08, AAAI Press, 1988