

Web 상에서 개인화된 상품 추천을 위한 Hybrid 추천 시스템에 관한 연구

손 창 환 (영남대학교 대학원 경영학과 박사수료)

김 기 수 (영남대학교 상경대학 경영학부 교수)

Abstract

인터넷의 성장은 고객에게 많은 혜택을 주었지만, 방대한 양의 정보는 오히려 장시간의 상품 탐색과 제품 선택을 어렵게 만들었다. 이에 따라, 정보의 양을 줄여 줄 수 있는 서비스를 고객들은 요구를 하기 시작하였고, 이에 따라 다양한 방법들이 고객에게 제시되어졌다.

제시되어진 방법 중의 하나가 개인화 추천 시스템이다. 추천 시스템은 고객의 취향과 관심에 적합한 상품을 추천 해 주는 서비스로서 상품 검색 노력을 줄여 주고, 고객의 취향에 적합한 제품을 제시 해 줌으로써 고객 충성도 제고에도 많은 도움을 주고 있다.

이러한 추천 시스템에서 가장 많이 사용되어지고 있는 기법은 협업 필터링이다. 협업 필터링은 협업에서도 유용한 기법으로 인정을 받았다. 하지만 희박성과 확장성이라는 문제점으로 인해 추천의 정확도가 다소 떨어진다는 것이 단점이다. 본 연구에서는 이러한 단점을 극복할 수 있는 방법으로써 Hybrid 협업 필터링 기법을 제시하고, 이를 토대로 추천 기법이 혼합되어진 Hybrid 추천 시스템에 대한 개념을 제시하고자 한다.

keywords: 협업 필터링, 아이템 기반 협업 필터링, 사용자 기반 협업 필터링, 지식기반 협업 필터링

I. 서론

인터넷의 급속한 성장과 더불어, 전자상거래도 빠르게 성장하였다. 이러한 인터넷의 빠른 성장으로 인하여 고객과 판매업자는 새로운 국면을 맞이하게 되어졌다. 판매업자는 경쟁이 심화되어져 살아남기가 어려워졌고, 고객은 방대한 양의 정보로 인하여 원하는 상품을 찾고 선택하기가 어려워졌다. 이러한 이유로, 일대일 마케팅(One-to-One Marketing)이나 고객관계관리(Customer Relationship Management) 등의 새로운 마케팅 전략이 요구되어졌다[Kim et al., 2002]

최근에는 새로운 마케팅 전략으로서 고객의 취향이나 관심에 적합한 상품을 추천하는 형태의 고객 맞춤 또는 개인화되어진 서비스가 널리 적용되어지고 있다. 고객의 취향이나 관심에 적합한 상품을 추천함으로써 고객의 상품 검색 노력을 줄여 줄 뿐만이 아니라, 적합한 상품 추천으로 인해 쇼핑몰 사이트에 대한 고객의 충성도 제고에도 도움을 주기 때문에 많이 이용되어지고 있다[Kim et al., 2002; Mild and Natter, 2002]

그래서, 고객의 취향에 맞는 상품을 추천하여 고객의 구매 결정을 도와주는 개인화 추천 시스템이 많이 이

용되어지고 있는데, Amazon.com, CDnow.com 등 해외의 우수한 전자상거래 사이트에 적용되고 있으며, Ringo 음악 추천이나 Bellcore 비디오 추천에도 이용되고 있다[Cyrus, 2003; Schafer, 1999].

현재 Amazon.com, CDnow.com, eBay, Levis 등의 사이트에서 운영하고 있는 추천 시스템에 가장 많이 적용되어진 추천 기법은 협업 필터링이다. 협업 필터링(Collaborative Filtering : CF)은 고객의 취향과 관심에 적합한 상품을 예측하고자 할 때 사용되어지는 정보 필터링 기법이다[Mukund, 2004; Schafer 1999]

이 기법은 고객의 과거 기록을 토대로 고객에게 적합한 상품을 예측하고자 하는 방법으로써 현재 가장 많이 사용되어지고 가장 유용한 기법으로 인정을 받았음에도 불구하고 다음과 같은 문제점이 존재한다[Cyrus, 2003; 김재경, 2005; 박지선, 2002], .

첫째, 시스템 확장성(Scalability) 문제이다. 고객과 상품의 수가 증가함에 따라 이웃 고객군을 찾기 위한 연산량이 기하급수적으로 늘어날 수 밖에 없기 때문에 실시간으로 추천을 목적으로 하는 상품추천시스템에서는 심각한 시스템 확장성 문제에 직면하게 된다. 둘째, 데이터의 희박성(Sparsity) 문제이다. 웹사이트에서 판매되는 상품의 수가 기하급수적으로 증가함에 따라 고객의 선호도가 입력되지 않은 상품의 개수가 상대적으로 많아짐으로 인해 이웃 고객군을 형성하는 과정에서 매우 적은 수의 평가 데이터만을 사용함으로써 유사도 측정에서 신뢰성이 떨어지고, 이는 결국 상품 추천의 질을 떨어뜨리는 요인으로 작용한다.

이와 같은 문제점을 보완하기 위해 본 연구에서는 사용자 기반 협업 필터링(User-based Collaborative Filtering)과 아이템 기반 협업 필터링(Item-based Collaborative Filtering)을 결합한 하이브리드 협업 필터링 기법(Hybrid Collaborative Filtering Techniques)과 지식기반기법인 사례기반추론 기법을 혼합한 하이브리드 추천 시스템(Hybrid Recommendation System)에 대한 방법론을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 협업 필터링 기법에 관한 관련 연구를 살펴보도록 하고, 3장에서는 Hybrid 협업 필터링의 구현 방법에 대해서 설명한다. 4장에서는 Hybrid 추천 시스템 구성에 대해 설명하고 5장에서 결론을 맺는다.

II. 관련 연구

2.1 추천 시스템의 추천 기법

추천 시스템은 추천 과정이 시작되기 전에 추천시스템 내에 존재하는 background data, 추천을 하기 위해 고객에 의해 입력되어지는 input data, 그리고 추천하기 위해 background data와 input data를 조합하는 알고리즘 등으로 형성되어진 것으로 고려할 경우, 5가지 유형의 추천 기법으로 분류 되어질 수 있으며 <표 1>과 같다[Robin, 2002]

Robin(2002)은 상기 기법 중에서 지식기반(Knowledge-based) 기법이 다른 기법들의 단점을 보완하는데 가장 좋은 방법이며, 특히 협업 필터링과 지식기반기법을 혼합하여 사용함으로써 협업 필터링의 문제점을 해결할 수 있음을 밝혀냈다.

<표 1> 추천 시스템의 추천 기법

기법	Background	Input	Prcess
Collaborative	Ratings from U of items in I	Ratings from u of items in I	Identify users in U similar to u, and extrapolate from their ratings of i.
Content-based	Features of items in I	u's ratings of items in I	Generate a classifier that fits u's rating behavior and use it on i
Demographic	Demographic information about U and their ratings of items in I	Demographic information about u	Identify users that are demographically similar to u, and extrapolate from their ratings of i
Utility-based	Feature of items in I	A utility function over items in I that describes u's preferences	Apply the function to the items and determine i's rank
Knowledge-based	Feature of items in I. Knowledge of how these items meet a user's needs	A description of u's needs or interests	infer a match between i and u's need

2.2 속성 기반 상품 계층도

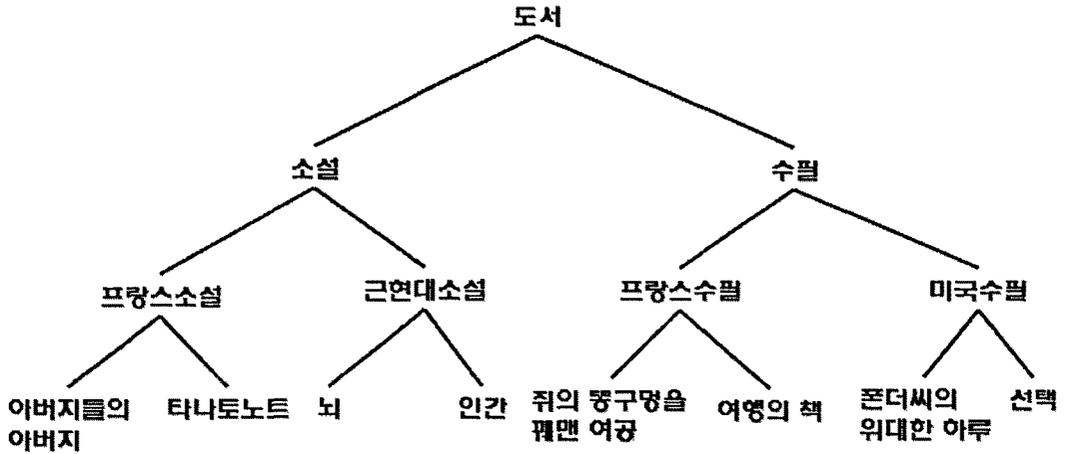
계층도는 개체 상호간의 연관성을 쉽게 표현할 수 있는 방법이며, 개체간의 공통적인 속성이나 유사성을 알기 쉽게 해주는 특성을 지니고 있다.

예를 들어, 고객 홍길동 군은 프랑스 소설 중에서 “아버지들의 아버지”와 “타나토노트”를 구매하고, 고객 김영희 양은 근현대 소설 중에서 “뇌”와 “인간”를 구매하고 고객 김철수 군은 프랑스 수필 중에서 “취의 퐁구멍을 깨맨 여공”과 “여행의 책”을 구매하였다고 하였을 경우, 도서 상품 계층도를 이용하면 각 상품간 유사성으로 인하여 고객간의 유사성도 쉽게 파악할 수 있다는 것이다[Prasanna, 2003].

고객간의 유사도를 파악하기 위해 Cosine 함수를 사용하여 보면 고객 “홍길동”, “김영희”, “김철수” 간의 유사도는 0으로 나타날 것이다. 고객들이 교차 구매한 상품이 없기에 유사도는 0으로 나타나는 것이다.

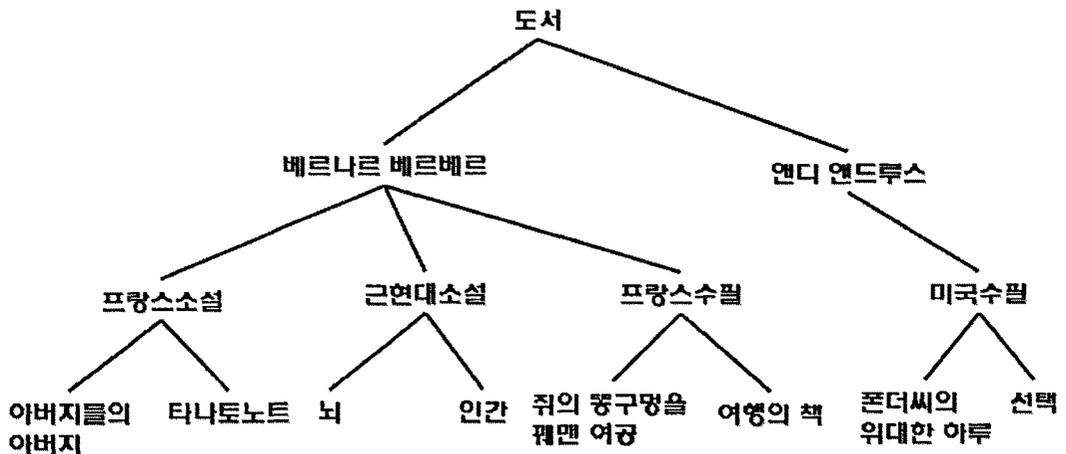
그러나 <그림 1>과 같은 도서 상품 계층도를 이용하게 되면 고객 “홍길동”과 “김영희”는 소설을 좋아하는 고객이라는 것과 고객 “홍길동”과 “김철수”는 성향이 유사하지 않은 고객이라는 사실을 쉽게 파악할 수 있게 되어진다.[Prasanna, 2003].

<그림 1> 도서 상품 계층도



만약, 고객 고길동 군이 프랑스 소설 중에서 “아버지들의 아버지”와 프랑스 수필 중에서 “쥐의 똥구멍을 껌맨 여공”을 구매하였다고 할 경우, 고객 고길동군이 고객 홍길동, 김철수 중에서 어떤 고객과 유사한 성향을 가지고 있는지를 상품 계층도로 파악할 수 없지만, 계층도 Level을 “지은이”라는 상품의 속성으로 구성한다면 파악할 수 있다. <그림 2>를 이용하면 홍길동 고객, 김영희 고객, 김철수 고객이 베르나르 베르베르 저자의 책을 선호하는 것으로 파악되어진다.. 이때 상품 속성 분류는 마케팅 전문가에 의해 이루어질 필요가 있다.

<그림 2> 속성 기반 상품 계층도



2.3 고객 프로파일링

인터넷 쇼핑몰에서의 고객 쇼핑 패턴은 상품탐색, 상세보기, 장바구니 담기, 구매 등의 4가지의 연속적인 과정으로 인식되어진다[Kim et al., 2002; Lee et al, 2001]

<표 2> 구매 패턴

단계	내용
Product Impression (탐색)	관심 있는 상품 탐색
Clickthrough (상세보기)	관심 있는 상품별 페이지 방문
basket placement (장바구니 담기)	관심 있는 상품을 장바구니에 담기
Purchase (구매)	관심 있는 상품 구매

각 단계에서 발생되어지는 데이터는 로그파일에 저장되어지는데, 탐색 단계의 경우는 쇼핑몰 사이트에서 제공하는 검색엔진을 이용하여 탐색이 이루어질 수도 있기 때문에 로그 파일에는 탐색 과정에 대한 데이터가 저장되지 않는다.

그래서 탐색 과정을 제외한 3가지 과정의 데이터를 이용하여 고객의 선호도가 계산되어진다. 이때 이루어지는 계산은 각 과정의 발생 횟수로 설정하여 선호도를 계산하는 것이 일반적인 형태인데, 김재경(2005)의 연구에서 제안되어진 선호도 계산방법에서도 상세보기의 횟수, 장바구니 담기 횟수, 구매 횟수를 이용하여 선호도를 계산하였다.

이와 같이 선호도를 계산하여 이를 토대로 고객 프로파일을 형성하고자 할 때에, Jonathan(2004)은 시간이 흐르면서 고객의 선호도가 변화되어진다는 것을 고려해 볼 필요가 있다고 하였다.

즉, 시시각각 급속하게 변화하는 고객의 선호도를 대변하기 위해서는 시간의 흐름에 따라 자료를 변형하고 이를 분석해 볼 필요성이 있다는 것이다.

예를 들어, 홍길동 군이 소나타III 자동차에 관심이 있어, 상품 페이지에 자주 방문하다가 삼성에서 SM5를 출시하였다는 것을 알게 되었다. 이후 소나타III 자동차 상품 페이지에는 더 이상 방문하지 않고 SM5 자동차 페이지에만 계속적으로 방문하였다.

이때, 소나타III 자동차 상품 페이지 방문횟수가 100이고, SM5 자동차 상품 페이지 방문횟수가 50이면 고객이 선호하는 자동차는 소나타III 자동차로 인식되어질 수 있다. 또한 소나타III 자동차 상품 페이지 방문 횟수보다 낮은 상품은 무조건 고객 선호도가 낮은 상품으로 인식되어지는 오류가 발생할 수 있다.

<그림 3> 홍길동의 방문횟수

2003년 3월 방문횟수

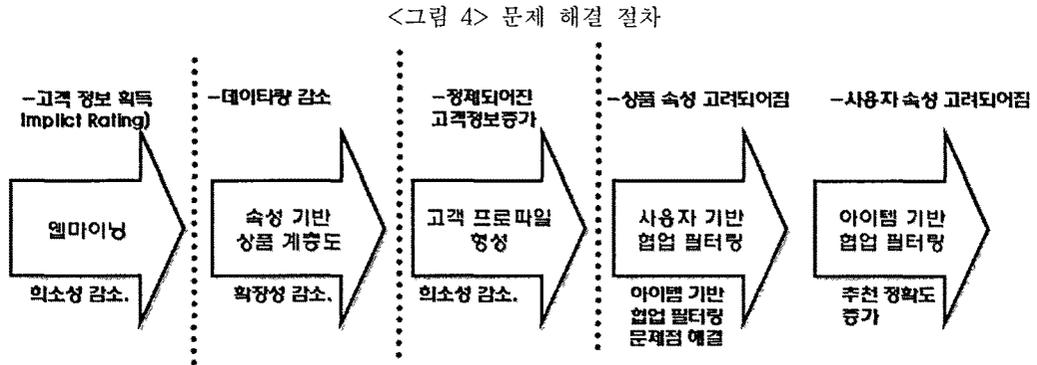
	소나타 III	SM 5
홍길동	100	0

2004년 3월 방문횟수

	소나타 III	SM 5
홍길동	100	50

III. Hybrid 협업 필터링

협업 필터링은 가장 강력하고 유용한 추천 기법으로 학계와 실무에서도 인정을 받아왔지만, 희박성과 확장성이라는 문제점으로 인하여 추천의 정확도가 다소 떨어진다는 것이다. 그래서 본 연구에서는 희박성과 확장성이라는 문제점을 해결하기 위한 방법론을 제시하고자 한다. 본 연구에서 제시하고자 하는 방법론의 전체적인 구성은 <그림 4>와 같다.



3.1 웹 마이닝

고객 프로파일은 상품에 대한 고객의 선호도에 의해 형성되어지는데, 고객의 선호도는 Implicit Rating과 Explicit Rating을 통하여 파악할 수 있다.

Explicit Rating은 주어진 항목에 대하여 고객이 평가를 함으로써 선호도를 파악할 수 있지만, Implicit Rating은 고객의 행동을 조사함으로써 선호도를 파악할 수 있다.

Implicit Rating은 고객이 상품을 구매함에 따라 발생되어지는 자료와 쇼핑몰 사이트 내에서의 쇼핑 행위 정보를 토대로 파악되어진다. 이때 고객이 쇼핑몰 내에서의 행동은 주로 로그 파일로 저장되어지는데, 로그 파일 내에는 날짜, 시간, 사용자 IP주소, 사용자이름, 접근방식, URI 스템, URI 쿼리, HTTP 상태, 쿠키, 참조페이지 등 고객에 대한 추가정보가 담겨지게 된다.

즉, 로그 파일에 저장되어 있는 정보는 고객 분석을 할 때 간과해서는 안 될 매우 중요한 정보이며, 고객의 성향이나 상품 선호도 분석에 있어 핵심적인 기초정보로 활용할 수 있다.

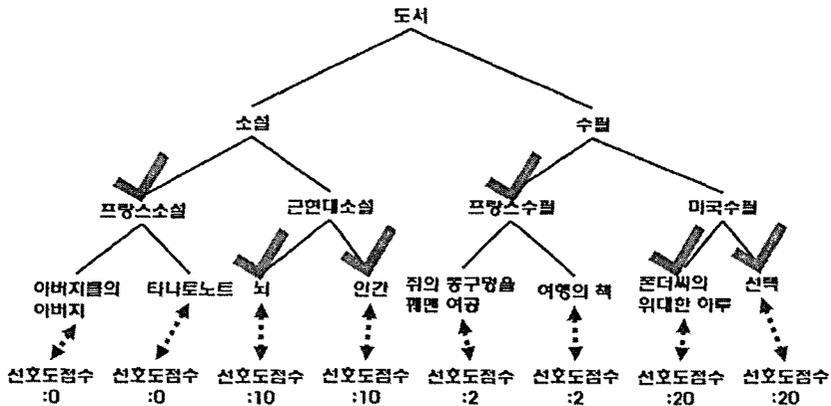
3.2 속성 기반 상품 계층의 형성

일부 선행연구에서는 일반적인 분류를 토대로 형성된 계층도를 이용하여 연구가 이루어졌지만, 이회정 외(2004)의 연구에서는 상품의 유형에 따른 분류보다는 상품의 속성에 따른 분류를 하는 것이 추천의 정확성을 높여 준다는 것을 밝혔다. 또한 Adomavicius et al(2001)의 연구에서는 속성 기반 계층도를 이용하여 제품군을 형성하는 것이 효과적이라는 것을 제시하고 있다.

협업 필터링의 희박성 문제점을 해결하기 위해서는 상품 속성에 따라 상품군을 분류하고 이를 토대로 속성 기반 상품 계층도를 형성할 필요성이 있다.

그리고 상품 계층 결정은 전체 고객 프로파일을 토대로 파악되어진 각 상품의 종합 선호도 점수로 이루어진다. 낮은 선호도를 가진 상품이 많은 경우에는 상위 상품군으로 선정되어지고, 선호도가 높은 상품이 많은 경우에는 개별 상품으로 계층이 결정되어진다. 쇼핑몰 이용자의 특성에 따라 상품군 계층도가 달라질 수 있다.

<그림 6> 상품 계층 결정



3.3 고객 프로파일 형성

웹 마이닝을 이용하여 고객에 대한 더 많은 정보를 획득하여 이를 토대로 고객의 프로파일을 형성하는 것이 중요하다. 고객 프로파일은 각 상품에 대한 고객의 선호도를 나타내는 것이기 때문에 중요하다. 또한, 협업 필터링의 문제점인 희박성을 해결하기 위해서라도 고객에 대한 정확하면서도 많은 정보를 획득하여 고객 프로파일을 형성할 필요가 있다.

이외에도 웹 마이닝을 이용하여 획득되어진 많은 자료를 적절하게 변형하고 정제하여 이를 토대로 고객 프로파일을 형성하는 것도 중요한 과정이라 할 수 있다.

그래서 본 연구에서는 Kim et al(2002), Lee et al(2001)의 연구에서 제시한 것과 같이, 인터넷 쇼핑몰에서의 고객 쇼핑 패턴을 4단계로 설정을 하고, 4단계 중에서 웹 로그에 기록되지 않는 탐색 단계를 제외한 나머지 단계(상세보기, 장바구니 담기, 구매)에서 발생되어지는 자료만을 적절한 정보로 변형하여 이를 토대로 고객 프로파일을 형성하고자 한다.

이외에도 고객의 프로파일을 형성하는데 있어서 추가적으로 고려해야 할 사항이 있는데, 고객의 선호도는 시간이 흐르면서 변화되어진다는 것이다.

이러한 요인으로 인하여, 본 연구에서는 쇼핑 4단계 중에서 상세보기 단계만 시간이 흐름에 따라 변화할 것이라는 가정하에 상세보기 횟수를 적절하게 변경하기 위해 평활법 중 이동평균법을 적용한다

<그림 8>은 3개월 이동평균으로 상세보기 횟수를 수정한 결과이다. 이때, 이동평균법의 기간은 평균 구매기간 또는 평균 방문기간을 파악하여 적용하면 된다.

<그림 6> 이동평균법을 이용한 상세보기 횟수

홍길종의 상세보기 횟수 / 일별

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
SM5	2	3	5	2	1	4	5	6	2	3	7	1

홍길종의 상세보기 횟수 / 일별 - 3개월 이동평균

	1월	2월	3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
SM5	-	3.3	3.3	2.6	2.3	3.3	5	4.3	3.6	4	3.6	-

조정되어진 상세보기 횟수, 구매 횟수, 장바구니 담기 횟수를 식 (1)을 이용하여 정규화 하도록 한다. 이후, 이를 토대로 최종 고객 프로파일을 형성하도록 한다.

$$P_{ij} = \frac{AV_{ij}}{\sum_{j=1} AV_{ij}} + \frac{B_{ij}}{\sum_{j=1} B_{ij}} + \frac{P_{ij}}{\sum_{j=1} P_{ij}} \quad (1)$$

식 (1)은 i 고객에 대한 선호도 점수를 정규화한 것이다. AV_{ij} 는 i 고객의 상품 j 에 대해 조정된 방문 횟수, B_{ij} 는 i 고객의 상품 j 에 대해 장바구니에 담은 횟수, P_{ij} 는 i 고객의 상품 j 에 대해 구매 횟수를 의미한다. 식 (1)을 이용하여 고객의 프로파일을 형성하면 <그림 9>와 같다.

<그림 7> 고객 프로파일 형성의 예

조정된 상세보기 횟수

$i \backslash j$	에어로션	손크림	케어편 에어	립글로즈	머드코백	에센스 에어플	폼글렌징	메이크업 베이스	에어왁스	립크로션
홍길종	10	12	8	1	4	3	2	1	1	1
홍길순	1	6	8	3	2	1	1	3	6	2
홍길동	2	3	5	2	1	4	5	6	2	3

장바구니 담기 횟수

$i \backslash j$	에어로션	손크림	케어편 에어	립글로즈	머드코백	에센스 에어플	폼글렌징	메이크업 베이스	에어왁스	립크로션
홍길종	4	3	5	0	1	0	3	0	0	0
홍길순	2	4	7	1	1	1	1	3	4	5
홍길동	2	3	5	2	1	4	5	6	2	3

구매향수

i \ j	에어로션	손크림	케어컨 헤어	립글로즈	머드코백	에센스 헤어플	점글렌징	메이크업 베이스	에어웍스	립크로션
홍길동	3	2	2	0	0	0	1	0	0	0
홍길순	1	2	3	0	1	1	1	1	1	2
홍길중	1	1	3	0	1	1	2	2	0	1

식 (1)에 의해 형성된 고객 프로파일

i \ j	에어로션	손크림	케어컨 헤어	립글로즈	머드코백	에센스 헤어플	점글렌징	메이크업 베이스	에어웍스	립크로션
홍길동	0.63	0.75	0.78	0.14	0.15	0.09	0.38	0.04	0.03	0.03
홍길순	0.15	0.46	0.66	0.18	0.18	0.16	0.17	0.29	0.34	0.42
홍길중	0.14	0.29	0.51	0.15	0.19	0.32	0.49	0.49	0.18	0.23

3.4 사용자 기반 협업 필터링

사용자 기반 협업 필터링의 첫 번째, 단계는 식 (2)를 이용하여 각 사용자별로 다른 고객과의 유사도를 구하는 것인데, 각 사용자별로 유사도를 구하여 이를 토대로 이웃 고객 집단을 형성하는 것이다.

$$r_{ij} = \frac{Cov(i, j)}{\sigma_i \sigma_j} = \frac{\sum_k (S_{ik} - \bar{S}_i)(S_{jk} - \bar{S}_j)}{\sqrt{\sum_k (S_{ik} - \bar{S}_i)^2 \cdot \sum_k (S_{jk} - \bar{S}_j)^2}}, -1 \leq r_{ij} \leq 1 \quad (2)$$

S_{ik} 는 고객 i 가 상품 k 에 대해 평가한 점수이고, \bar{S}_i 는 고객 i 의 평가 점수의 평균이다. S_{jk} 는 고객 j 가 상품 k 에 대해 평가한 점수이고, \bar{S}_j 는 고객 j 의 평가 점수의 평균이다.

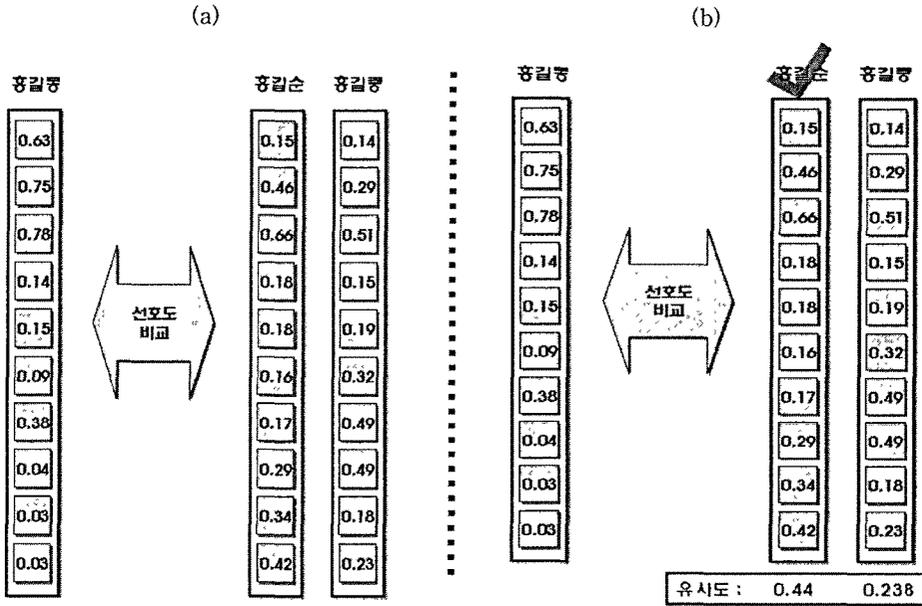
r_{ij} 는 고객 i 와 j 의 상관관계를 나타내며, 상관계수 r_{ij} 는 두 고객의 선호도가 유사한 경우에는 1에 가까운 값을 가지게 되고, 상반된 선호도를 갖는 경우는 -1에 가까운 값을 가지게 된다.

<그림 8> 최종 고객 프로파일

식 (1)에 의해 형성된 고객 프로파일

i \ j	에어로션	손크림	케어컨 헤어	립글로즈	머드코백	에센스 헤어플	점글렌징	메이크업 베이스	에어웍스	립크로션
홍길동	0.63	0.75	0.78	0.14	0.15	0.09	0.38	0.04	0.03	0.03
홍길순	0.15	0.46	0.66	0.18	0.18	0.16	0.17	0.29	0.34	0.42
홍길중	0.14	0.29	0.51	0.15	0.19	0.32	0.49	0.49	0.18	0.23

<그림 9> 선호도 비교



식 (2)에 의해서 고객간의 선호도를 파악한 결과, 홍길순은 0.44, 홍길중은 0.238로서 유사한 선호도를 가진 것으로 나타났다. 홍길중에 비해 유사도가 높은 홍길순만을 이웃 고객 집단으로 분류를 하였다.

두 번째 단계는, 이웃 고객 집단이 형성되어지면 식 (3)을 이용하여 특정 상품에 대한 선호도를 예측한다.

$$P_{ik} = \overline{S}_i + \frac{\sum_{j \in Raters(k)} (S_{jk} - \overline{S}_j) r_{ij}}{\sum_{j \in Raters(k)} |r_{ij}|} \quad (3)$$

고객 i 의 상품 k 에 대한 선호도 점수를 예측하기 위해서 상품 k 를 평가한 고객들 j 의 평가점수와 이들 고객들과 고객 i 의 상관계수를 이용하여 계산되어진다. 이때, $Raters(k)$ 는 상품 k 를 평가한 고객의 집합을 의미한다.

상기 사례에서의 홍길동 고객의 상품 선호도 예측값은 4.1로 나타났다. 이와 같이 식 (3)을 이용하여 얻은 예측값은 이웃고객에 따라 다른 값들이 나오게 되어지는데, 이 값의 크기에 따라서 추천 상품이 결정되어진다.

하지만 사용자 기반 필터링에서 사용한 피어슨 상관계수는 선호도를 표시하지 않은 상품인 경우에는 유사도를 측정할 수 없기 때문에 아이템 기반 협업 필터링을 재적용하도록 한다. 즉, 상품에 대한 고려가 되지 않았다는 것이다. 상품간의 유사도가 전혀 고려되지 않음으로 인하여 추천의 정확도가 다소 떨어지게 된 것이다.

3.5 아이템 기반 협업 필터링

선호도에 따라 고객을 분류하고 이를 토대로 추천을 하는 사용자 기반 협업 필터링 방식과 달리 아이템 기반 협업 필터링은 상품간의 유사성을 파악한 후 이를 토대로 고객이 기존에 선호한다고 평가한 상품과 유사한 상품을 찾아서 고객에게 추천해 주는 기법이다.

사용자 기반 협업 필터링을 적용한 데이터를 이용함에 따라 아이템 기반 협업 필터링의 문제점을 해결 할 수 있다. 아이템 기반 협업 필터링의 문제점은 상품들의 유사도를 계산하기 위하여 고객들간의 유사도가 전혀 고려되지 않는 한계점을 해결할 수 있다는 것이다.

아이템 기반 협업 필터링을 하기 위해서는 상품 x, y 간의 정확한 유사도 측정을 위해 먼저 상품 x, y 에 선호도를 모두 입력한 고객들을 별도로 추출하여야 한다. 이때 추출되어진 고객들의 집합을 U 라고 설정한다.

집합 U 에 속한 고객들이 상품 x, y 에 입력한 선호도에 의하여 식 (4)와 같이 x, y 의 유사도 $Sim(x, y)$ 가 계산된다.

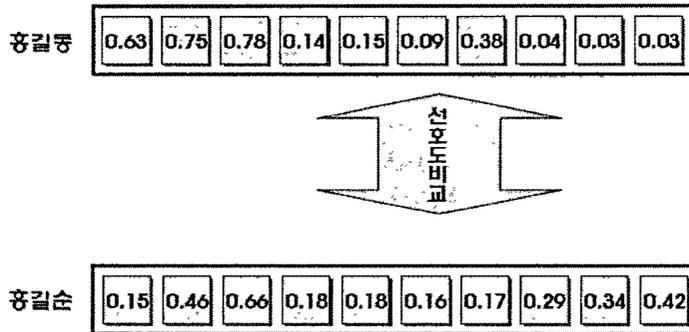
$$Sim(x, y) = \frac{\sum_{u \in U} (R_{u,x} - \overline{R_u})(R_{u,y} - \overline{R_u})}{\sqrt{\sum_{u \in U} (R_{u,x} - \overline{R_u})^2 \sum_{u \in U} (R_{u,y} - \overline{R_u})^2}} \quad (4)$$

, $-1 \leq Sim(x, y) \leq 1$

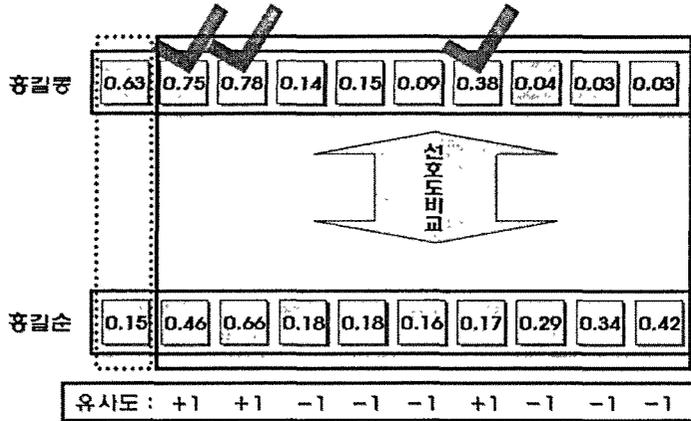
위의 식에서 $R_{u,x}$ 는 고객 u 가 상품 x 에 대해 평가한 선호도를 나타내며, $\overline{R_u}$ 는 고객 u 가 평가한 선호도의 평균을 의미한다.

사용자 기반 협업 필터링은 행렬의 열따라 계산되지만, 아이템 기반 협업 필터링은 행렬의 열따라 계산되어지기 때문에, 피어슨 상관 계수보다는 조정된 Cosine을 이용하여 유사도를 계산한다[Bardul, 2001] 식 (4)를 이용하여 상품에 대한 선호도를 먼저 파악하도록 한다.

<그림 10> 선호도 비교



<그림 11> 유사 상품



상품 x 와 유사한 상품에 대해 고객이 평가한 선호도의 합을 계산하여서 고객 u 가 상품 x 에 대해 가지는 선호도를 예측한다. 이때, 상품 x 와 y 사이의 유사도에 따라서 각각의 선호도 값이 가중치 된다.

식 (5)에 의해 고객 u 의 상품 x 에 대한 선호도 예측 값을 계산 할 수 있다. 여기에서 고객 u 의 x 와 비슷한 상품들에 대한 선호 경향에 따라 상품 x 에 대한 u 의 선호도 값을 예측한다.

$$P_{u,x} = \frac{\sum_{all\ similar\ items, N} (s_{i,N} * \overline{R_{u,N}})}{\sum_{all\ similar\ items, N} (|s_{i,N}|)} \quad (5)$$

예측값은 각 상품에 따라 다른 값들이 나오게 되어서, 이 값의 크기에 따라 최종적인 N개의 추천 상품이 형성되어진다.

IV. Hybrid 추천 시스템

Robin(2002)은 5가지 추천 기법 중에서 지식기반(Knowledge-based) 기법이 다른 기법들의 단점을 보완하는데 가장 좋은 방법이며, 특히 협업 필터링과 지식기반기법을 혼합하여 사용함으로써 협업 필터링의 문제점을 해결할 수 있음을 밝혀냈다.

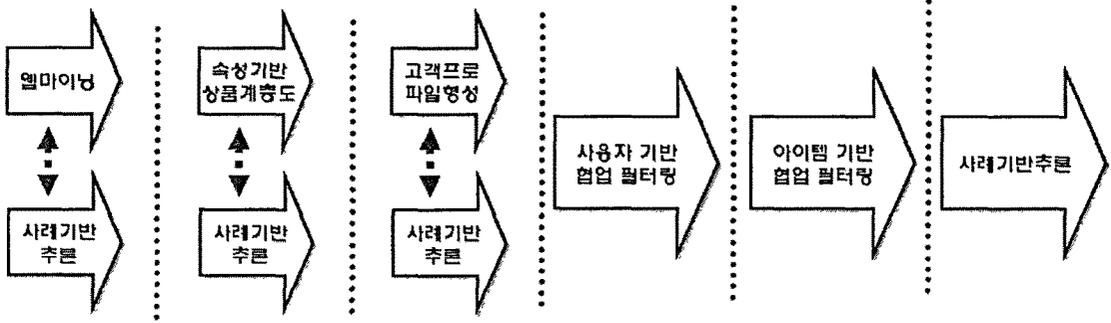
Robin(2002)의 연구에 따라 본 연구에서는 Hybrid 협업 필터링과 지식 기반 추천 기법을 혼합한 추천 시스템을 제안하고자 한다. 기존 Hybrid 협업 필터링의 기본적인 절차나 처리과정은 동일하지만, 지식기반 추천기법으로 인식되는 사례기반추론을 혼합하여 사용하는 것이 다소 틀리다고 할 수 있다.

사례기반추론을 이용함으로써 신규고객이나 신상품에 따른 문제점을 해결할 수 있다. 여기서 문제점이라는

것은 신상품이나 신규고객의 경우에는 선호도가 입력되지 않음으로 인해 추천과정에서 배제가 되어진다는 것이다.

사례베이스에 포함되어져 있는 기존 사례에 의해서 신규고객에게 상품을 추천하거나 신상품을 추천에 포함할 수 있는 것이 가능해지는 것이다.

본 연구에서 제시하고자 하는 추천 시스템의 전체적인 구성은 다음과 같다.

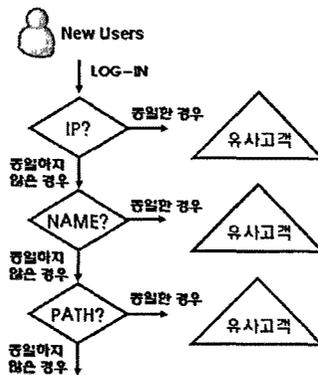


웹마이닝을 이용하여 획득되어진 정보는 2차 가공되어 사례베이스에 저장되어진다. 그리고 계층도를 형성하기 위해 분류한 속성에 대한 정보와 고객 프로파일에 관한 정보들도 2차 가공되어 사례베이스에 저장되어지게 된다. 추가적으로 고객에 대한 분석이 가능해 질 수도 있고, 기존 고객의 유형에 맞추어 신상품이나 신규고객에 적절한 상품을 추천해 줄 수 있는 시스템이라 할 수 있다.

추가적으로 웹마이닝을 이용하는 방법을 달리 해 봄으로써 추가적인 정보 획득이 가능해진다. 일반적으로 웹 로그 분석을 통해 얻을 수 있는 정보는 1차원적인 정보뿐이다. 하지만 사이트 내에 존재하는 고객 프로파일 또는 구매 기록 등을 추가하여 복합적으로 웹 로그 분석을 실시함으로써 추가적인 정보 획득이 가능할 것이다. 이러한 분석은 고객 특성별로 사이트 이용도, 상품 구매에 관련되어진 고객 행동 패턴 등에 대한 폭넓은 분석이 가능하도록 할 것이다.

추가적인 분석이 가능하며, 기존에 파악 할 수 없었던 고객의 선호도에 대한 추가적인 정보 획득이 가능하기 때문에 웹 마이닝을 적극 이용할 필요성이 있다.

<그림 5> 로그파일을 이용하여 유사고객 파악(신규고객의 경우)



이외에도, Hybrid 추천 시스템을 이용하여 추천 상품을 형성하고자 할 경우에는 인구통계학적 요인을 고려할 필요성도 있다.

<표 3>은 한국 광고단체 연합회/한국 인터넷 마케팅 협의회가 주관하는 KNP (Korea Netizen Profile) Survey의 2003년, 2004년 조사 보고서 중에서 인터넷을 통해 구매한 제품/서비스에 관한 설문조사의 결과를 정리한 것이다. 이 결과에 의하면 성별에 따라 구매하는 제품이나 서비스가 다르다는 것을 인식할 수 있다.

추천의 정확도를 높이기 위하여 남/녀 데이터를 별도로 하여 추천 기법을 적용할 필요성이 있다는 것이며, 이외에도 연령이나 지역에 따라 추천기법을 달리 해보는 것도 유의한 결과를 도출할 수 있다.

<표 3> KNP Survey 2003, 2004 연차보고서

	남 -2004	녀 -2004	남 -2003	녀 -2003
기타			11.7	19.8
꽃배달			4.4	2.1
아동유아용품	6.2	11.3	0.3	9.4
개인잡화	39.9	55	42.5	65.7
의류	30.2	48.9		
컴퓨터관련	45	14.1	38.7	11.1
가전전자제품	43	28.3	49.3	28.6
비디오테잎	9.3	11.9	40.9	49.4
도서	40	48.1		
예약	12.5	12.4	12.9	13.5

출처 : <http://www.advertising.co.kr/blog/com/index.vw> [광고정보센터]

V. 결론

인터넷상에서 운영되어지는 쇼핑물의 수가 점차적으로 증가하고 있고, 쇼핑물에서 취급되어지는 상품의 수도 점차적으로 증가하고 있다. 이와 더불어, 상품에 대해 제공되어지는 정보의 양도 많아져서, 선택의 폭을 넓혀주었다. 하지만, 방대한 양의 정보는 오히려 장시간의 상품 탐색과 제품 선택을 어렵게 만들었다.

이에 따라, 정보의 양을 줄여 줄 수 있는 서비스를 고객들은 요구를 하기 시작하였고, 이에 따라 다양한 방법들이 고객에게 제시되어졌다.

이러한 이유로 제시되어진 방법 중의 하나가 개인화 추천 시스템이다. 추천 시스템은 고객의 취향과 관심에 적합한 상품을 추천 해 주는 서비스로서 상품 검색 노력을 줄여 주고, 고객의 취향에 적합한 제품을 제시 해 줌으로써 고객 충성도 제고에도 많은 도움을 주고 있다.

Amazon.com, CDnow.com 등 해외의 우수한 전자상거래 사이트에 적용되고 있으며, Ringo 음악 추천이나 Bellcore 비디오 추천에도 이용되고 있다[Cyrus, 2003; Schafer, 1999]. 현재 Amazon.com, CDnow.com, eBay, Levis 등의 사이트에서 운영하고 있는 추천 시스템에 가장 많이 적용되어진 추천 기법은 협업 필터링이다. 협업 필터링(Collaborative Filtering : CF)은 고객의 취향과 관심에 적합한 상품을 예측하고자 할 때 사용되어지는 정보 필터링 기법이다[Mukund, 2004; Schafer 1999]

이 기법은 고객의 과거 기록을 토대로 고객에게 적합한 상품을 예측하고자 하는 방법으로써 현재 가장 많이 사용되어지고 가장 유용한 기법으로 인정을 받았음에도 불구하고 희박성과 확장성이라는 문제점으로 인하여 추천의 정확도가 떨어진다는 것이 큰 단점이다.

본 연구에서는 이러한 단점을 극복할 수 있는 방법으로써 웹마이닝과 속성 기반 상품 계층도, Hybrid 협업 필터링 기법을 확장성과 희박성 문제점 해결 방안으로 제시했지만, 제시한 방법론에 대해 검증은 하지 못하였다.

웹 로그 정보와 방문횟수와 같은 고객에 관한 실제 데이터를 구하기가 어려워 검증을 실시하지 못하였지만, 협업 필터링의 확장성과 희박성의 문제점을 해결 시킬 수 있는 방안을 제시하는 것으로 연구의 의의를 두고자 한다.

그리고 추가적으로 제시한 Hybrid 추천시스템은 웹 로그를 복합적으로 분석함으로써 고객에 대한 폭넓은 분석이 가능하도록 할 것으로 기대되어지며, 이에 대한 연구를 추후 진행할 예정이며, 차후의 연구에서는 제시한 방법론을 검증하기 위해 추천 시스템을 실제로 구축하여 운영을 할 예정이다.

또한 추천 시스템의 기본적인 문제점인 신규고객과 신상품에 대한 문제 해결 방안을 제안하였는데, 향후 신규고객과 신상품의 문제추천과정에서 포함되지 못함으로써 발생하는 문제에 대한 연구가 향후 이루어질 필요가 있다.

참 고 문 헌

- [1] 김영설, 김병천, 윤범주, “개선된 추천시스템을 이용한 전자상거래시스템 설계 및 구현”, 정보처리학회논문지, 제9-D권, 제2호, 2002.4.
- [2] 김재경, 안도현, 조윤희, “개인별 상품 추천 시스템, WebCF-PT:웹마이닝과 상품계층도를 이용한 협업필터링”, 경영정보학연구, 제15권, 제1호, 2005년, 3월, pp63-79.
- [3] 박지선, 김택헌, 류영석, 양성봉, “추천시스템을 위한 2-way 협동적 필터링 방법을 이용한 예측 알고리즘”, 정보과학회지 논문지, 29권, 9,10호, 2002년 10월, pp669-675.
- [4] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms”, WWW10, 2001, Hong Kong.. 285- 295
- [5] Cyrus Shahabi Yi-shin Chen, “An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback?”, *Distributed and Parallel Databases*, 14, 2003, pp173-192.

- [6] Gediminas Adomavicus, Alexander Tuzhilin, "Expert-Driven Validation of Rule-Based User Models in Personalization Applications", *Data Mining and Knowledge Discovery*, 5, 2001, pp33-58.
- [7] Jae Kyeong Kim, Yoon Ho Cho, Woo Ju Kim, Je Ran Kim, Ji Hae Suh, "A personalized recommendation procedure for Internet shopping support", *Electronic Commerce Research and Applications 1*, 2002, pp301-313.
- [8] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News", *COMMUNICATIONS OF THE ACM*, vol 40, no 3, 1997, pp77-87
- [9] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, John T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", *ACM Transactions on Information Systems*, Vol. 22, No. 1, January 2004, pp5-53.
- [10] Juhnyoung LEE, Mark Podlaseck, Edith Schonberg, Robert Hoch, "Visualization and Analysis of Clickstream Data of Online Stores for Understanding Web Merchandising", *Data Mining and Knowledge Discovery*, 5, 2001, pp59-84.
- [11] Mukund Deshpande and George Karypis, "Item-Based Top-N Recommendation Algorithms", *ACM Transactions on Information Systems*, Vol. 22, No. 1, January 2004, pp 143-177.
- [12] Paul Resnick, Hal R. Varian, "Recommender systems", *Communications of the ACM*, Volume 40, Issue 3, March 1997, pp 56-58.
- [13] Prasanna Ganesan, Hector Garcia-Molina, and Jennifer Widom, "Exploiting Hierarchical Domain Structure to Compute Similarity", *ACM Transactions on Information Systems*, Vol. 21, No. 1, January 2003, pp 64-93.
- [14] Robin Burke, "Hybrid Recommender systems : Survey and Experiments", *user modeling and user-adapted interaction*, 12, 2002, pp331-370.
- [15] Schafer J. Ben , Joseph Konstan, John Riedl, Recommender Systems in E-Commerce, In Proceedings of ACM-E-Commerce. ACM. New York. 1999.