

# Recommender System based on Product Taxonomy and User's Tendency

Heonsang Lim\* · Yong Soo Kim\*\*†

\*Samsung Electronics

\*\*Department of Industrial and Management Engineering, Kyonggi University

## 상품구조 및 사용자 경향성에 기반한 추천 시스템

임헌상\* · 김용수\*\*†

\*삼성전자

\*\*경기대학교 산업경영공학과

In this study, a novel and flexible recommender system was developed, based on product taxonomy and usage patterns of users. The proposed system consists of the following four steps : (i) estimation of the product-preference matrix, (ii) construction of the product-preference matrix, (iii) estimation of the popularity and similarity levels for sought-after products, and (iv) recommendation of a products for the user. The product-preference matrix for each user is estimated through a linear combination of clicks, basket placements, and purchase statuses. Then the preference matrix of a particular genre is constructed by computing the ratios of the number of clicks, basket placements, and purchases of a product with respect to the total. The popularity and similarity levels of a user's clicked product are estimated with an entropy index. Based on this information, collaborative and content-based filtering is used to recommend a product to the user. To assess the effectiveness of the proposed approach, an empirical study was conducted by constructing an experimental e-commerce site. Our results clearly showed that the proposed hybrid method is superior to conventional methods.

Keywords : Recommender System, Collaborative Filtering, Content based Filtering, Product Taxonomy

## 1. 서 론

### 1.1 연구의 필요성

최근 인터넷 서비스 및 모바일 기기의 발달에 따라 언제 어디서든 네트워크에 접속할 수 있는 시대가 도래되었다. 또한, 수많은 콘텐츠와 서비스가 범람하고 있으므로, 사용자 입장에서는 자신이 원하는 콘텐츠를 원하는 시점에 전달받는 것은 매우 필요한 기능이다. 즉, 사용자가 선

호하리라 판단되는 콘텐츠 및 상품을 적절하게 추천하는 시스템이 범용적으로 활용되고 있으며, 이러한 추천 시스템을 활용하여 개인화 서비스가 이루어지고 있다[6, 8].

추천 시스템은 각 사용자에게 각 사용자가 선호하거나 구매할만한 콘텐츠 및 상품을 추천함으로써, 사용자의 선택을 돕는 시스템이다. 특히, 매우 방대한 양의 상품을 가진 쇼핑몰 사이트에서 사용자가 모든 상품을 살펴보기 어려울 때, 더욱 높은 효용성을 지닌다.

아마존으로 대표되는 기존의 쇼핑몰 추천 시스템에서는 구매이력을 바탕으로 각 사용자 별로 맞춤형 정보를 제공함으로써 유용한 상품을 구매할 수 있도록 정보를 제공하고 있다. 즉, 사용자의 선호도를 구매 여부를 대상

Received 21 April 2013; Finally Revised 24 May 2013;

Accepted 22 June 2013

† Corresponding Author : kimys@kgu.ac.kr

으로 한 것이다. 그러나, 최근의 연구에서는 사용자의 구매 이력뿐만 아니라 클릭 및 검색 정보를 바탕으로 사용자의 선호도를 추정하는 연구가 진행되어 왔다[4, 5, 6]. 그러나, 많은 전자상거래 사이트의 추천시스템에서는 상품의 구조에 대한 고려가 부족할 뿐만 아니라 사용자의 선호 경향성과는 무관하게 동일한 알고리즘을 적용하는 연구들이 주로 이루어졌다.

따라서, 본 연구에서는 더욱 예측력이 높은 추천 시스템을 개발하기 위해, 전자 상거래 사이트의 구조에 기반하여 사용자의 선호 경향성에 따라 각각 다른 알고리즘을 적용하는 방법론을 제안하고자 한다.

## 1.2 문헌 연구

현재까지 추천시스템에 관한 많은 연구가 이루어졌으며, 그 중 가장 보편적인 것이 협업적 필터링(collaborative filtering)과 내용기반 필터링(content based filtering) 기법이다. 협업적 필터링이란, 각 사용자에 대해 해당 사용자와 유사성이 가장 높은 이웃집합(neighbor)을 찾아낸 다음, 이웃집합이 선호하는 상품 및 콘텐츠를 해당 사용자에게 추천하는 것이다[8, 11]. 비슷한 선호도를 지닌 이웃집합을 찾을 때는 Pearson correlation coefficient[11], constrained Pearson correlation coefficient[14], Jaccard coefficient[7] 및 cosine vector[7] 등이 사용되며, 이웃집합을 구한 다음 각각의 가중평균으로 각 사용자가 접하지 못한 상품 또는 콘텐츠에 대한 선호도를 예측하게 된다. 이 때 각 유사도 측도에 따라 이웃집합이 다르게 형성될 수 있으므로, 선호도 점수도 상이하게 나타날 수 있다[17].

그러나, 협업적 필터링의 경우 데이터가 많지 않은 경우에는 각 사용자와 선호 정보가 유사한 이웃집합을 찾기 어려우므로, 유용하지 못할 뿐만 아니라, 사용자 수 및 상품 수가 증가함에 따라 연산속도도 기하급수적으로 증가하는 경향이 있다[7, 8]. 또한 협업적 필터링은 사용자가 직접 평가한 데이터에 대해서는 효과적이거나, 전자상거래 사이트에서의 구매 여부와 같은 이진 데이터에 대해서는 예측력이 저하되는 문제점이 있다[4, 7].

한편, 내용기반 필터링 기법은 사용자간의 선호 정보 대신 각 사용자 별 특정 콘텐츠 또는 상품의 속성을 학습하는 것이다. 본 방법론은 정보 추출 분야에서 발전된 것으로서, 사용자와 콘텐츠간의 유사도를 계산하여, 그 결과를 순위화 하는 과정이다[9]. 따라서, 각 상품 및 콘텐츠의 속성이 제대로 정의되어 있어야 하며, 새로운 콘텐츠별로 속성에 대한 값의 입력이 요구된다. 따라서, 영상, 이미지, 문서 등과 같은 콘텐츠의 경우 그 내용을 정의하는 것이 매우 어려우므로 범용성이 부족하다[15, 16]. 또한, 내용기반 필터링 기법은 사용자가 과거에 선

호했던 속성을 지닌 상품 및 콘텐츠가 계속해서 추천되는 단점이 있다[6, 7].

앞서 언급한 협업적 필터링 기법과 내용기반 필터링 기법의 단점을 극복하기 위해 두 기법을 접목한 하이브리드 기법에 대한 연구도 이루어졌다[1, 2]. 대부분의 하이브리드 기법은 협업적 필터링에 기반한 채 내용기반 필터링의 개념을 일부 도입한 것으로서, 내용기반 필터링 기법에서 요구하는 각 아이템별 사용자 프로파일을 요구하고 있다[2]. 따라서, 각 콘텐츠에 대한 정의가 이루어져야 하므로 범용적으로 적용하기 어려운 단점이 있다. 반면, 협업적 필터링 기법과 내용기반 필터링 기법의 결과를 적절하게 반영하는 혼합형 추천시스템에 대한 연구도 수행되었다[5]. Kim[5]이 제안한 혼합형 문서 추천 시스템에서는 사용자의 이용 패턴에 따라 각 사용자 별로 협업적 필터링과 내용기반 필터링 기법 결과를 각각 다른 비율로 적용하여 추천 목록을 제공하는 연구가 이루어졌다[5].

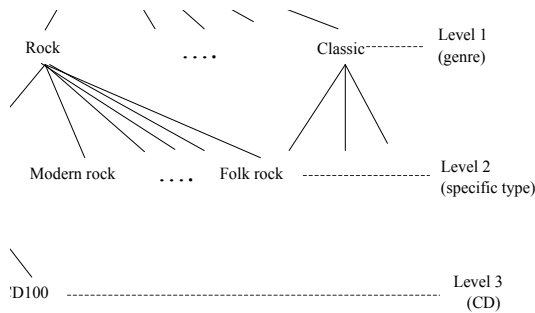
본 연구의 대상인 전자 상거래 사이트의 많은 연구에서는 주로 구매데이터만을 활용함으로써 데이터의 밀도가 매우 낮아 높은 성능을 발휘하지 못하였다. 이에 반해 Kim et al.[7]은 사용자의 클릭 정보, 장바구니 담기 정보, 구매 정보는 물론 상품 페이지 머문 시간, 선호 장르 비율, 선호 세부장르 비율 등을 활용하여 높은 성능을 발휘하는 협업적 필터링 기법을 제안하였다[7]. 그러나, 해당 기법은 매우 많은 연산량을 지니므로 실제 구현에 어려움이 존재하였다. 따라서, 이를 해결하기 위해 데이터를 종류별로 분할하여 연산속도를 향상시키는 한편 예측력은 유사하거나 다소 높게 나타나는 연구가 이루어졌다[6].

본 연구에서는 전자 상거래 사이트를 대상으로 상품 구조에 따라 데이터를 분할한 다음, 사용자의 클릭 경향성을 고려하여 협업적 필터링 기법과 내용기반 필터링 기법을 혼합하는 새로운 방법론을 제안하고자 한다. 그 다음, 동일한 데이터에 대해 전체 데이터를 대상으로 협업적 필터링을 적용한 연구결과[7]와 상품 구조에 따른 분할된 데이터를 대상으로 협업적 필터링을 적용한 연구결과[6]와 비교 평가하였다.

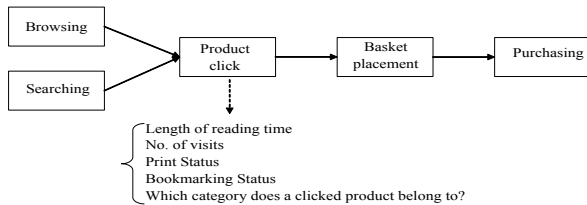
Kim et al.[7] 연구에서는 사용자의 행동 및 탐색 패턴 데이터를 바탕으로 선호도를 추정하였으나, 사용자의 경향성을 반영하지 못하였다. 따라서, Kim[5]에서는 예측력을 향상시키기 위해 사용자 경향성을 반영한 문서 추천시스템을 개발하였고, Kim[6]에서는 데이터를 분할하여 적용함으로써, 연산량을 감소시키는 추천시스템을 개발하였다. 따라서, 본 연구에서는 분할된 데이터를 대상으로 사용자의 상품 선호도 및 사용 경향성까지 반영하여 예측력이 높은 전자상거래 추천시스템을 개발하고자 한다.

## 2. 전자상거래 사이트에서 수집한 데이터 형태

전자상거래 사이트에서는 주로 상품의 구조가 정형화 되어 있으며, 각 수준별로 상품을 진열하거나 관리한다. 아래의 <Figure 1>은 음반판매 사이트의 상품구조를 나타낸 것이며, <Figure 2>는 사용자가 음반판매 사이트에 로그인(log-in)에서부터 음반 구매에 이르기까지 수집 가능한 데이터를 표현한 것이다.



<Figure 1> Product Taxonomy of CD E-commerce Site[7]



<Figure 2> Possible Actions That Can be Taken by Users in E-commerce Sites and Possible Data that Can be Obtained From Such Actions [7]

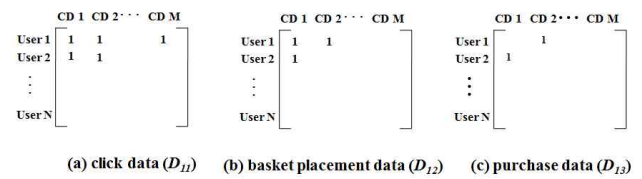
## 3. 개발된 추천 시스템

본 연구에서는 크게 다음과 같이 세 가지로 단계로 구성된다. 첫째, 선호도 추정 과정으로서 사용자 행동 유형별 데이터로부터 각 사용자의 음반 선호도, 세부장르 선호도 및 장르 선호도를 추정한다. 둘째, 각 사용자의 이용 행태를 파악하는 과정으로서, 각 사용자의 인기 상품 선호도 및 특정 장르 선호도를 추정한다. 셋째, 각 사용자에게 어떠한 알고리즘을 어떠한 비율로 적용할지 결정하는 과정으로서, 협업적 필터링 기법과 내용기반 필터링 기법을 혼합하여 사용자에게 추천을 하게 된다.

### 3.1 선호도 추정 과정

본 연구에서는 ‘구매여부’, ‘장바구니 담기 여부’ 및 ‘클릭

여부’의 데이터로부터 각각 ‘각 상품별 구매 선호도’, ‘각 상품별 장바구니 선호도’, ‘각 상품별 클릭 선호도’ 행렬을 산정하였다. 그 다음, 각각의 가중합을 통해 ‘각 상품별 선호도’, ‘각 세부장르별 선호도’, ‘각 장르별 선호도’ 행렬을 각각 구하였다. 그 외의 변수인 ‘클릭한 상품을 읽은 시간’, ‘북마킹 여부’, ‘프린팅 여부’ 등은 기존의 연구에서 선호도 추정시 중요 변수가 아니었을 뿐 아니라, 빠른 연산 처리를 위하여 본 연구에서는 제외하였다. 각 상품별로 사용자의 이용 단계에 따른 행렬은 아래의 <Figure 3>과 같다.



<Figure 3> Sample User-product Preference Matrix[5]

그 다음, 위의 각 행렬( $D_{11}$ ,  $D_{12}$ ,  $D_{13}$ )에 대해 가중합을 구함으로써, ‘사용자-상품 선호도 행렬’을 다음과 같이 산정한다.

$$D_1 = \alpha_1 D_{11} + \beta_1 D_{12} + \gamma_1 D_{13}$$

이 때,  $0 \leq \alpha_1 \leq 1$ ,  $0 \leq \beta_1 \leq 1$ ,  $0 \leq \gamma_1 \leq 1$ ,  $\alpha_1 + \beta_1 + \gamma_1 = 1$ 인 관계를 갖는다.

다음 단계에서는 장르 및 세부장르별 선호도 데이터를 구성한다. 즉, 위의 절차를 각 상품에 대해서가 아니라, 각 장르별 또는 세부장르별로 ‘사용자-세부장르 선호도 행렬’과 ‘사용자-장르 선호도 행렬’을 각각 구한다. 이때, 각 상품에 해당하는 CD의 클릭 빈도수는 단순합을 통해 구한다(<Figure 4> 참조).

$D_{21}$ ,  $D_{22}$ ,  $D_{23}$ 과  $D_{31}$ ,  $D_{32}$ ,  $D_{33}$ 로부터 각각 ‘사용자-세부장르 선호도 행렬’ ( $D_2$ )과 ‘사용자-장르 선호도 행렬’ ( $D_3$ )을 구하며, 각 원소의 계산식은 다음과 같다. 본 연구에서는  $d_{mn-ij}$ 은  $D_{mn}$  행렬의  $i$ 번째 행,  $j$ 번째 열의 원소값을 의미하며,  $d_{m-ij}$ 은  $D_m$  행렬의  $i$ 번째 행,  $j$ 번째 열의 원소값을 나타낸다

$$d_{2-ij} = \frac{\alpha_2 d_{21-ij} + \beta_2 d_{22-ij} + \gamma_2 d_{23-ij}}{\sum_{j=1} d_{2-ij}}$$

$$d_{3-ij} = \frac{\alpha_3 d_{31-ij} + \beta_3 d_{32-ij} + \gamma_3 d_{33-ij}}{\sum_{j=1} d_{3-ij}}$$

예를 들어, 사용자 1이 장르 1에 해당하는 음반을 3

회 클릭하고, 2회 장바구니에 담은 후, 1회 구매까지 이루어졌다고 가정할 때, 사용자 1의 장르 1에 대한 선호도는 다음과 같이 계산된다(<Figure 4> 참조).

$$d_{3-11} = \frac{\alpha_3 d_{31-11} + \beta_3 d_{32-11} + \gamma_3 d_{33-11}}{\sum_{j=1}^3 d_{3-1j}}$$

$$= \frac{0.2 \times 3 + 0.3 \times 2 + 0.5 \times 1}{(0.2 \times 3 + 0.3 \times 2 + 0.5 \times 1) + (0.2 \times 5 + 0.3 \times 4 + 0.5 \times 1) + (0.2 \times 6 + 0.3 \times 1 + 0.5 \times 0)}$$

$$= 0.288 (\text{if } \alpha = 0.2, \beta = 0.3, \gamma = 0.5)$$

### 3.2 사용자의 경향성 분석 과정

각 사용자의 이용 경향성을 파악함으로써, 각 사용자에게 협업적 필터링이 적합할 것인지 내용기반 필터링 기법이 적합할 것인지 결정하기 위한 기초 데이터를 생성하는 단계이다. 사용자의 클릭 패턴으로부터 추정된 각 음반에 대한 선호도, 세부장르에 대한 선호도, 장르에 대한 선호도, 인기 콘텐츠 선호도 및 장르 유사도, 선호 가수 유사도 등의 측도를 고려하게 된다.

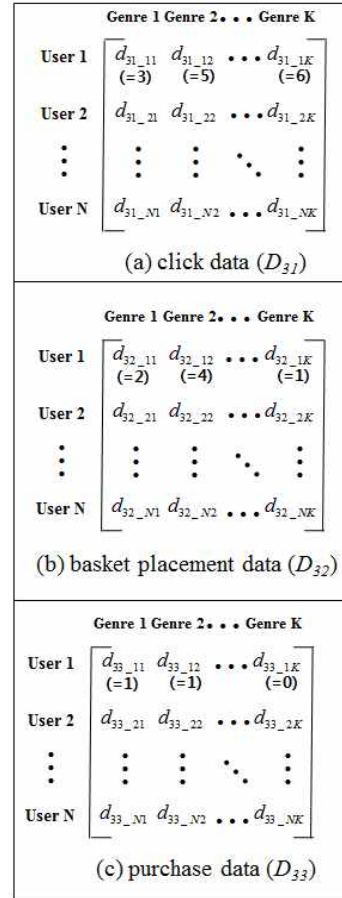
유사도 또는 다양도 등을 측정하기 위해 본 연구에서는 정보 추출분야에서 사용되는 엔트로피 지수를 사용하였다[10]. 엔트로피 지수는 예전부터 물리학, 생물학, 심리학, 경제학은 물론 정보 추출분야에서 다양도 측도로써 널리 이용되고 있으며[3] 다음과 같이 계산된다[13].

$$Entropy(u_a) = - \sum_{i=1}^c \left\{ \frac{up_i}{N_a} \times \log_2 \left( \frac{up_i}{N_a} \right) \right\}$$

여기서  $u_a$ 는 사용자  $a$ 를,  $c$ 는 클릭한 장르의 개수를,  $up_i$ 는 사용자  $a$ 의 장르  $i$ 에 대한 선호도를,  $N_a$ 는 사용자가 클릭한 모든 장르 선호도 합을 의미한다.

예를 들어, 어떤 고객의 장르 엔트로피 지수가 낮다면, 해당 사용자는 유사한 장르의 음반을 선호하는 것이고, 엔트로피 지수가 높다면 상이한 장르의 음악을 선호하는 것을 의미한다. 만일, 어떤 사용자의 엔트로피 지수가 낮고 인기도 지수가 높다면, 그 사용자는 대중적인 음악을 선호하며 유사한 장르의 음악을 주로 선호하는 것을 알 수 있다. 이러한 사용자에게는 내용기반 필터링 기법을 적용함으로써, 사용자가 선호하는 유사 장르의 음악을 추천하거나, 인기도가 높은 음악을 추천하는 것이 합당할 것이다.

본 연구에서는 사용자의 장르 엔트로피 지수와 각 사용자의 인기 음반 선호도를 각각 구하였다. 각 상품의 인기도는 한 번 이상 클릭된 음반의 클릭 수를 내림차순으로



<Figure 4> Example of a User-genre Preference Matrix[5]

정렬한 다음, 각각의 차지하는 비율과 누적 비율을 계산하였다. 그리고 1에서 각 누적 비율을 차감한 것을 인기도로 결정하였다(<Table 1> 참조). 그 다음, 각 사용자가 클릭한 음반의 인기도 점수의 평균을 구하여, 이를 사용자의 인기 상품 선호도로 결정한 것이다(<Table 2> 참조).

<Table 1> Example: Calculation of CD Popularity Levels

CD id	Number of Clicks	Occupation Ratio	Cumulative Occupation Ratio	CD Popularity Level
CD 2	10	0.50	0.50	0.50(= 1-0.50)
CD 3	6	0.30	0.80	0.20(= 1-0.80)
CD 1	3	0.15	0.95	0.05(= 1-0.95)
CD 4	1	0.05	1.00	0.00(= 1-1.00)

<Table 2> Example: Calculation of User Popularity Levels

User ID	Clicked CDs	User Popularity Level
User 1	CD 2, CD3	0.35(= (0.50 + 0.20)/2)
User 2	CD2, CD1, CD 4	0.35(= (0.50 + 0.20 + 0.05)/3)
User 3	CD 1	0.05

예를 들어, 사용자 1이 음반 2와 음반 3을 클릭하였다면, 두 음반의 인기도 점수의 평균인 0.35가 사용자 1의 인기상품 선호도로 결정하는 것이다.

또한, 엔트로피 지수는 특정 장르에 대한 집중도를 파악하기 위한 것으로서, 본 연구에서는 과 ‘사용자-장르 선호도 행렬’ ( $D_3$ )을 대상으로 엔트로피 지수를 산정하였다. 이 때, 상위 25%를 ‘high’로, 하위 25%를 ‘low’로, 나머지를 ‘medium’으로 결정하였다. 또한, 인기도는 각 사용자의 인기도 점수가 상위 20%면 ‘high’로, 나머지는 ‘low’로 설정하였다(<Table 3> 참조).

<Table 3> Reflection Ratio of Recommendation Methods

User Group	Genre Entropy Index	Popularity Level	Reflection Ratio		
			CF	CBF	Popular CDs
Group 1	Low	High	0.3	0.3	0.4
Group 2	Low	Low	0.3	0.5	0.2
Group 3	Medium	Low	0.3	0.5	0.2
Group 4	Medium	High	0.3	0.2	0.5
Group 5	High	High	0.6	0.1	0.3
Group 6	High	Low	0.7	0.3	0.0

### 3.3 추천 알고리즘 적용 비율 결정 과정

본 절에서는 각 사용자 그룹에 따라 협업적 필터링, 내용기반 필터링, 인기 상품 추천 반영 비율을 결정하는 과정이다. 즉, 각 기법별로 Top-N개를 추출한 다음, 각 기법에서 몇 개씩의 아이템을 추천할 것인지 결정하는 것이다. 예를 들어, 엔트로피 지수가 낮고 인기상품 선호도가 높은 그룹에 대해 인기상품 10개 및 협업적 필터링과 내용기반 필터링 기법에 의해 각 10개씩 추출한 다음, 그것을 2개, 3개, 5개씩 추천할 것인지 아니면 5개, 0개, 5개를 추천할 것인지 결정한다.

이 때, 인기 상품은 전체 사용자가 가장 많이 클릭한 상품 Top-N개를 추출하였다. 또한, 협업적 필터링 기법은 가장 전통적인 메모리기반 협업적 필터링 기법을 사용하되, 이웃집합 형성과정에서는 ‘사용자-세부장르 선호도행렬 ( $D_2$ )’ 사용하였고, 예측치는 ‘사용자-상품 선호도행렬( $D_1$ )’을 사용하였다. 이웃집합 형성과정에서  $D_2$  행렬을 사용한 것은 연산속도를 빠르게 하고, 전통적 협업적 필터링 기법의 단점인 희소성 문제(sparseness problem)를 극복하기 위함이다. 한편, 내용기반 필터링 기법은 사용자의 아티스트 선호도, 장르 선호도, 세부장르 선호도를 바탕으로 cosine 측도를 활용하여 가장 근접한 상품을 추천하였다. cosine 측도는 다음과 같이 계산된다.

$$\cos(\vec{U}, \vec{T}) = \frac{\vec{U} \cdot \vec{T}}{\|\vec{U}\| \times \|\vec{T}\|}$$

여기서,  $\vec{U}$ 는 각 사용자의 장르(세부 장르 또는 아티스트) 선호도를,  $\vec{T}$ 는 각 음반의 해당 장르(세부 장르 또는 아티스트)를 의미한다.

본 연구에서는 <Table 3>에서 보는 바와 같이 총 6개의 그룹에 대해 full factorial로 세 그룹의 반영 비율을 결정하였다. 전체 데이터를 활용하여 인기상품 목록과 각 사용자별 추천 목록을 협업적 필터링 및 내용기반 필터링을 활용하여 각각 구하였다. 그 다음, 각 사용자 그룹별로 각 기법에서 추출된 목록의 반영 비율을 full factorial을 통해 추천한 것이다. 이 때, ‘사용자-상품 선호도행렬( $D_1$ )’ 중 10%의 데이터를 무작위로 삭제한 다음, 삭제된 영역의 예측값과 실측값의 평균최소제곱합을 최소로 하는 비율을 다음과 같이 결정하였다.

## 4. 실험 평가

이와 같이 산출된 반영 비율을 활용하여 기존의 연구 방법인 전체 데이터에 동일한 알고리즘을 적용한 기존 연구와의 차이를 분석할 필요가 있다. 따라서, 본 연구에서는 새로 개발된 방법론과 사용자의 동일한 클릭 데이터에 대해 협업적 필터링 기법만을 적용한 기존의 방법, 분할된 데이터에 대해 협업적 필터링 기법만을 적용한 방법론을 비교평가 하였다.

각 방법론의 성능을 평가하기 위해서는 다음의 방법을 사용한다. 실제 구매가 이루어진 1의 값을 갖는 원소 중 10%를 무작위로 삭제한다. 즉, 삭제된 원소는 구매는 물론, 클릭도 이루어지지 않은 상품으로 간주한다. 각 추천 알고리즘에 의해 결측치에 예측값을 구한 다음, 예측값의 내림차순으로 Top-N 목록을 작성한다. 본 연구에서는  $N = 10, 20, 30$ 으로 정하여 비교평가 하였다. 마지막으로, Top-N 목록이 삭제된 상품을 얼마나 포함되는지 분석함으로써 성능을 평가한다.

성능 평가척도로는 정보추출분야에서 널리 활용되는 ‘recall’과 ‘precision’의 조화평균인 F1 측도를 활용한다. ‘recall’은 삭제된 상품 중 얼마나 많은 상품이 추천목록에 포함되었는지 평가하는 척도이고, ‘precision’은 추천 목록에 포함된 상품 중 삭제된 상품이 얼마나 많이 포함되었는지 평가하는 척도이다[5, 6, 7, 12].

$$recall = \frac{\sum_{i \in A} |H_i \cap Top\_N_i|}{\sum_{i \in A} |H_i|}$$

$$precision = \frac{\sum_{i \in A} |H_i \cap Top\_N_i|}{N \times |A|}$$

여기서,  $H_i$ 는 '고객  $i$ 의 숨겨진 상품'을,  $N$ 은 '각 고객별로 추천된 상품의 수'를,  $Top-N$ 은 '고객  $i$ 에게 추천된  $Top-N$  목록'을,  $A$ 는 '하나 이상의 숨겨진 상품을 지닌 고객들'을 의미한다. 'recall'과 'precision'은 'recall'이 증가하면 'precision'이 감소하는 경향이, 'precision'이 증가하면 'recall'이 감소하는 경향이 있으므로, 두 측도의 조화 평균인 F1 값이 널리 사용된다[6, 10].

$$F1 = \frac{2 \times recall \times precision}{recall + precision}$$

본 연구에서도 F1 값을 이용하여 추천 시스템의 성능을 평가하였다. 아래의 <Table 4>는 실제 구매 데이터의 10%를 무작위로 삭제한 다음, Top-10, Top-20, Top-30으로 변화하며 F1 값을 비교한 것이다.

<Table 4> Experimental Results

N	Proposed Method	Conventional Methods	
		CF based on Entire Data Set [6]	CF based on Divided Data Set [5]
10	0.060	0.028	0.045
20	0.038	0.024	0.027
30	0.022	0.020	0.022

## 5. 결론 및 추후 연구 방향

본 연구에서 제안한 방법이 기존의 연구보다 우수하거나 유사한 성능을 나타냈다. 추천 목록 개수가 10, 20인 경우에는 본 연구의 방법론이 기존의 연구들보다 우수하게 나타났으나, 30인 경우에는 동일하게 나타났다. 즉, 추천 목록 개수가 적을수록 높은 예측력을 나타내었다. 추천 시스템은 적은 수의 사용자 선호 아이템을 추천하는 것이므로, 이것은 의미 있는 결과라고 할 수 있다.

또한, 본 연구는 음반판매 전자상거래 사이트를 대상으로 연구를 수행하였으나, 본 연구에서 수행한 방법론은 다른 전자상거래 사이트에서도 얼마든지 적용할 수 있다. 전자상거래 사이트는 상하구조를 지니고 있으므로, '장르', '세부장르', '상품' 대신 각 수준을 다른 명칭으로 대입할 수 있을 것이다. 뿐만 아니라, 각 상품이 갖는 속성도 음반이 지닌 속성(예시 : 장르, 아티스트, 세부 장르 등)을 각 영역에 맞게 정의할 수 있다면, 다른 분야에서도 충분히 적용 가능하다.

추후 연구에서는 본 연구에서 적용한 전통적 협업적 필터링 기법 대신 특이값 분해 기법(SVD : Singular Value Decomposition) 등과 같은 차원 감소 기법[10]을 적용하여 더욱 높은 예측력을 갖는 추천 시스템을 개발할 필요가 있다. 아울러, 추천 속도를 높일 수 있도록 연산량을 고려한 추천 시스템의 설계가 요구된다.

또한, 본 연구의 방법론을 전자상거래 사이트뿐만 아니라, 사용자가 직접 콘텐츠를 생산하는 UCC 사이트 및 고객간의 거래가 이루어지는 C2C(Customer to Customer) 서비스인 경매 사이트와 같은 다양한 영역에 적용한다면, 추천 시스템의 효용성을 더욱 높일 수 있으리라 판단된다.

## Acknowledgement

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MEST) (NO.2011-0014553).

## References

- [1] Bose, R., Advanced Analytics : Opportunities and Challenges, *Industrial Management and Data Systems*, 2009, Vol. 109, No. 2, p 155-172.
- [2] Choi, S.H., Jeong, Y.S., and Jeong, M.K., A Hybrid Recommendation Method with Reduced Data for Large-Scale Application, *IEEE Transactions on Systems, Man, and Cybernetics-Part C : Applications and Reviews*, 2010, Vol. 40, No. 5, p 557-566.
- [3] Curtrini, E., Using Entropy Measures to Disentangle Regional from National Localization Patterns, *Regional Science and Urban Economics*, 2009, Vol. 39, p 243-250.
- [4] Hayes, C., Cunningham, P., and Smyth, B., A Case-Based Reasoning view of Automated Collaborative Filtering, *Proceedings of the Fourth International Conference on Case-Based Reasoning*, 2001, p 243-248.
- [5] Kim, Y.S., Text Recommender System Using User's Usage Patterns, *Industrial Management and Data Systems*, 2011, Vol. 111, No. 2, p 282-297.
- [6] Kim, Y.S., Recommender System based on Product Taxonomy in E-Commerce Sites, *Journal of Information Science and Engineering*, 2013, Vol. 29, p 63-78.
- [7] Kim, Y.S., Yum, B.-J., Song, J., and Kim, S.M., Development of a Recommender System based on Navigational and Behavioral Patterns of Customers in E-Commerce

- Sites. *Expert Systems with Applications*, 2005, Vol. 28, No. 2, p 381-393.
- [8] Lawrence, R.D., Almasi, G.S., Korlyar, V., Viveros, M.S., and Duri, S.S., Personalization of Supermarket Product Recommendations, *Data Mining and Knowledge Discovery*, 2001, Vol. 5, No. 1, p 67-77.
- [9] Mooney, R.J. and Roy, I., Content-based Book Recommending using Learning for Text Categorization, *Proceedings of the 18<sup>th</sup> National Conference on Artificial Intelligence*, 2000, p 187-192.
- [10] Rennolls, K., Likelihood, Entropy and Species Diversity : Some Comparisons in a Sumatran Forest, *Proceedings of Forest Biometry, Modeling and Information Sciences*, 2001.
- [11] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedle, J., Grouplens : An Open Architecture for Collaborative Filtering of Netnews, *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 1994, p 175-186.
- [12] Sarwar, B., Karypis, G., Konstan, J.A., and Riedle, J., Analysis of Recommendation Algorithms for E-Commerce, *Proceedings of the ACM E-Commerce Conference*, 2000, p 158-167.
- [13] Shannon, C.E., A Mathematical Theory of Communication, *Bell System Journal*, Vol. 27, p 379-42.
- [14] Shardanand, U. and Maes, P., Social Information Filtering : Algorithms for Automating Word of Mouth, *Proceedings of Conference on Human Factors in Computing Systems*, 1995, p 210-217.
- [15] Shyu, M.-L., Chen, S.-C., Chen, M., Zhang, C., and Sarinnapakorn, K., Image Database Retrieval Utilizing Affinity Relationships, *Proceedings of the First ACM International Workshop on Multimedia Databases*, 1995, p 78-85.
- [16] Sun, J., Wang, Z., Yu, H., Nihino, F., Katusyama, Y., and Naoi, S., Effective Text Extraction and Recognition for WWW Images, *Proceedings of the 2003 ACM Symposium on Document Engineering*, 2003, p 115-117
- [17] Wei, C.-P., Yang, C.-H., and Hsiao, H.-W., A Collaborative Filtering-based Approach to Personalized Document Clustering, *Decision Support Systems*, 2008, Vol. 45, p 413-428.