추천 선행평가에 의한 마케팅 도메인 및 고객군 선정

윤찬식* · 이수원**

Selecting Marketing Domains and Customer Groups by Pre-evaluation on Recommendation

Chan-Sik Yune* · Soo-Won Lee**

요 약

협력적 추천 기법은 유사한 이웃의 선호도를 이용하여 고객에게 개인화된 아이템을 추천해 주는 방법으로 비교적 높은 정확도를 보이며 추천 시스템의 중심으로 연구되어져 왔다. 그러나, 지금까지의 추천 시스템은 도메인의 특성을 제대로 고려하지 못한 채 추천을 시행함으로써 특정 도메인에서 추천의 정확도가 떨어지는 문제점이 발생하였다. 이러한 문제점들을 보완하기 위하여 본 논문에서는 평균 고객 유사도, 평균 아이템 유사도, 밀집도 등의 추천 선행 평가 척도를 제안하고, 추천 선행평가 척도와 추천의 정확도와의 상관관계를 보이며, 이를 이용하여 짧은 수행시간 안에 추천 적용이 가능한 마케팅 도메인 및 고객군을 선정하는 방법을 제시한다.

Key words : 사용자기반 협력적 추천(User-based Collaborative Recommendation), 아이템기반 협력적 추천(Item-based Collaborative Recommendation), 선행평가(Pre-evaluation), 평균 사용자 유사도(Average User Similarity), 평균 아이템 유사도(Average Item Similarity), 피어슨 상관계수(Pearson Correlation Coefficient), 밀집도(Density), Support Filter, Monetary, Frequency

1. 서 론

인터넷 및 IT 기술의 발달하고 고객의 요구사항이 점차 다양화됨에 따라, 고객에 대한 정보를 효율적으로 관리하기 위한 데이터웨어하우징 및 유용한 정보의 발견을 위한 데이터마이닝에 대한 연구가 활발히 진행되고 있으며, 이를 기반으로 사용자의 기호에 맞는 개인화된 서비스 및 정보를 제공해 주는 추천시스템의 중요성이 대두되고 있다.

대부분의 추천 시스템은 협력적 추천(Collaborative Recommendation) (Sarwar et al, 2001), 내용기반 추천 (Content-based Recommendation) (Balabanovic et al, 1997), 연관규칙(Association Rule)을 이용한 추천(Lin et al, 1999), 군집화(Clustering)를 통한 추천(Sarwar et al, 2000), 인구통계학적(Demographic)(Michael, 1999) 추천 등의 추천기법을 사용하고 있다. 협력적 추천이란 추천하고자 하는 아이템에 대해 유사 사용자들의 선호도를 반영하여 추천하는 방법이며, 내용기반 추천은 유사 사용자를 고려하지 않고 아이템의 내용을 분석하여 새로운 아이템이 들어왔을 때 유사한 아이템을 추천하는 방법이다 (Balabanovic et al, 1997). 또한, 연관규칙을 통한 추천은 트랜잭션을 대상으로 항목간의 연관성을 발견하여 추천하는 방법이고, 군집화를 통한 추천은 사용자가 소속되는 군집의 다른 사용자들의 선호도를 통하여 개인화된 추천에 적용하는 방법이며, 인구통계학적 추천은 사용자들의 인구통계학적 정보를 이용하여 유사한 사용자 그룹 특성 및 대표 상품을 추출하고 그룹 특성과 비슷한 사용자에게 그룹의 대표상품을 추천하는 방법이다.

이와 같이 다양한 추천 기법들이 제안되어 왔으나, 지금까지의 추천 시스템은 대부분 도메인의 특성을 충분히 고려하지않고 추천 알고리즘을 적용하여 왔다. 즉, 추천을 적용하기에 앞서 도메인의 특성을 분석하여, 추천이 잘 적용될지 여부를 판단하지 않고 추천 기법을 적용함으로써, 추천의 정확도를 떨어뜨리는 결과가 발생하였다. 이러한 문제점을 개선하기 위하여 본 논문에서는 평균 고객 유사도, 평균 아이템 유사도, 밀집도 등의 추천 선행평가 척도를 제안하고, 추천 선행평가 척도와 추천의 정확도와의 상관관계를 보이며, 이를 이용하여 짧은 수행시간 안에 추천 적용이 가능한 마케팅 도메인 및 고객군을 선정하는 방법을 제시한다.

* 숭실대학교 컴퓨터학과 박사과정
** 숭실대학교 컴퓨터학부 부교수

본 논문은 모두 5장으로 구성되어 있다. 2장에서는 관련연구들을 소개하며, 3장에서는 본 논문에서 제안하는 마케팅 도메인 및 고객군 선정을 위한 선행 평가척도를 제안한다. 4장에서는 본 논문에서 제안하는 선행평가 방법을 실제 데이터에 적용한 실험결과를 분석하며, 마지막으로 5장에서는 본 논문에 대한 결론 및 향후 연구를 기술한다.

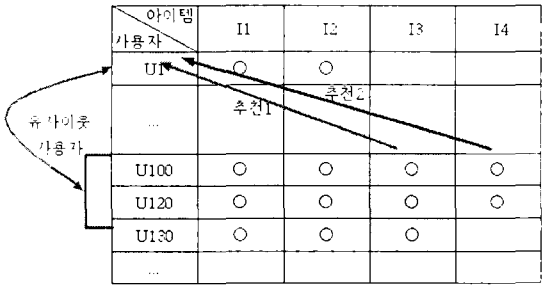
2. 관련연구

2.1 협력적 추천

협력적 추천이란 사회적 여과(Social Filtering)라고도 하며 유사한 기호를 가지는 다른 사람들의 선호도에 기반해서 추천하고자 하는 아이템을 여과한다(Shardanand et al, 1995). 협력적 추천 방법은 유사한 대상 선정 기준에 따라 크게 사용자 기반 협력적 추천과 아이템 기반 협력적 추천으로 나뉘어진다. 본 장에서는 사용자 기반 협력적 추천과 아이템 기반 협력적 추천에 대하여 기술하며, 유사도 산출, 추천 결과 예측, 추천 결과 평가와 관련된 기존 연구를 기술한다.

2.2 사용자기반 협력적 추천

사용자 기반 협력적 추천은 추천의 대상이 되는 사용자에 대하여 비슷한 선호 패턴을 갖는 유사 이웃 사용자들을 찾고, 이웃 사용자들이 선호하는 아이템들 중에서 아이템들을 선정하여 추천하는 기법이다 (Schafer et al, 2001).

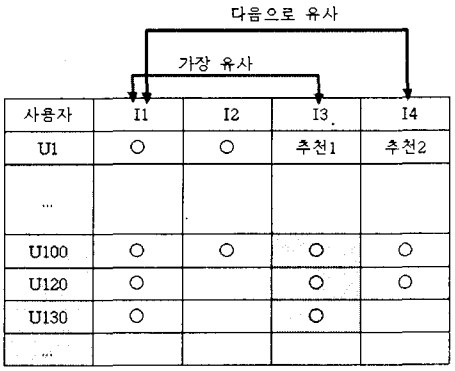


[그림 1] 사용자기반 협력적 추천

[그림 1]은 사용자 U1에게 U1과 유사한 이웃 사용자들이 선호하는 아이템을 추천하는 과정을 도해한 것이다. U1과 유사한 이웃으로 U100, U120, U130이 있고, 이들이 선호한 아이템 중에서 U1이 구매하지 않은 아이템은 I3와 I4가 있다. I3와 I4 아이템 중 I3는 세 명의 이웃이 선호를 표시하였으며, I4는 두 명의 이웃이 선호를 표시하였다. 따라서, 협력적 추천은 U1에게 U1의 이웃이 가장 선호한 I3를 먼저 추천하고 다음으로 I4를 추천한다

2.3 아이템기반 협력적 추천

아이템 기반 협력적 추천 기법은 사용자 기반 협력적 추천 기법이 항상 전체 고객을 대상으로 유사한 이웃을 찾는 행위를 반복적으로 수행하는 단점을 극복한 추천 기법으로, 추천의 대상이 되는 아이템과 유사한 이웃 아이템들을 찾고, 추천 대상 고객이 유사 이웃 아이템들에게 선호한 등급(Rating)에 관련 아이템의 유사정도를 가중치로 반영하는 추천 기법이다



[그림 2] 아이템기반 협력적 추천 기법

[그림 2]는 아이템 I1에 대하여 I3이 가장 유사하고 다음으로 I4가 유사할 때, 사용자 U1이 I1을 선호하면, 먼저 I1과 가장 유사한 I3을 추천한 후, I1과 다음으로 유사한 I4를 추천하는 하는 과정을 도해한 것이다. 그림에서처럼 아이템 기반 협력적 추천은 아이템간 유사도를 기반으로 새로운 사용자에게도 추천을 적용할 수 있는 이점을 갖고 있다.

2.4 유사도 산출

사용자 기반 협력적 추천에서는 유사도 또는 상관관계를 비교함에 있어서 대상이 사용자가 되어 서로 유사한 사용자들을 찾는 기법이며, 아이템 기반 협력적 추천에서는 대상이 아이템이 되어 서로 유사한 아이템을 찾는 기법이다.

대상간의 유사도 또는 상관관계를 산출하는 방법으로는 피어슨 상관계수(Pearson Correlation Coefficient), 코사인 계수(Cosine Coefficient), 유클리디언 거리계수(Euclidean Distance Coefficient) 등이 대표적이다. 피어슨 상관계수(Pearson Correlation Coefficient)(Resnick, 1994)를 사용한 유사도 산출은 식1과 같다.

$$S_{a,u} = \frac{\sum_{i=1}^{m}(r_{a,i}-\bar{r_a})*(r_{u,i}-\bar{r_u})}{\sqrt{\sum_{i=1}^{m}(r_{a,i}-\bar{r_a})^2}\sqrt{\sum_{i=1}^{m}(r_{u,i}-\bar{r_u})^2}} = \frac{\sum_{i=1}^{m}(r_{a,i}-\bar{r_a})*(r_{u,i}-\bar{r_u})}{\sigma_a * \sigma_u} \quad \text{식1}$$

식1은 사용자 기반 협력적 추천 기법에서 두 사용자 a와 u사이의 상관관계 $S_{a,u}$를 구하는 식으로, m은 아이템의 수이며, $r_{a,i}$는 사용자 a가 아이템 i에게 제공한 Rating을 나타내고, $\bar{R_a}$는 사용자 a의 모든 아이템 항목에 부여한 Rating의 평균을 나타내며, σ_a는 사용자 a가 모든 아이템 항목에 대해 부여한 Rating의 표준편차를 나타낸다.

아이템 기반 협력적 추천 기법에서 피어슨 상관계수의 사용 방법은 식2와 같으며, 식1에서 사용자에 해당하는 수식기호를 아이템으로 대치한다.

$$sim(i,j)=S_{i,j}=\frac{\sum_{u=1}^{U}(r_{u,i}-\bar{r})*(r_{u,j}-\bar{r})}{\sqrt{\sum_{u=1}^{U}(r_{u,i}-\bar{r})}\sqrt{\sum_{u=1}^{U}(r_{u,j}-\bar{r})}}=\frac{\sum_{u=1}^{U}(r_{u,i}-\bar{r})*(r_{u,j}-\bar{r})}{\sigma*\sigma}$$ 식2

식2는 아이템 기반 협력적 추천 기법에서의 아이템간의 유사도 $sim(i,j)$를 산출하는 식이다. $|U|$는 전체 고객의 수를 나타내며, $r_{u,i}$는 사용자 u가 아이템 i에 대해 부여한 Rating을 나타내고, $\bar{r}_i$는 사용자들이 아이템 i에 대해 부여한 Rating의 평균을 나타내며, σ_i는 아이템 i에 대하여 사용자들이 부여한 Rating의 표준편차를 나타낸다.

피어슨 상관계수의 값의 범위는 -1에서 1사이의 값을 가지며, 1에 가까울수록 유사도(Similarity)가 증가하고, -1에 가까울수록 비유사도(Dissimilarity)가 증가하며, 0일 때 서로 관계가 없다고 해석된다.

유사도 산출의 두 번째 방법은 코사인 계수(Cosine Coefficient)(Sarwar et al, 2001)에 의한 방법이다.

$$sim(i,j)=\cos(\vec{i},\vec{j})=\frac{\vec{i}\bullet\vec{i}}{\|\vec{i}\|\times\|\vec{j}\|}$$ 식3

식3은 벡터 모델로서 사용자간 또는 아이템간의 유사성을 구하는 식이다. i와 j는 고객기반 협력적 추천에서는 사용자에 해당하며, 아이템기반 협력적 추천에서는 아이템에 해당한다. 따라서, 식3은 두 고객 또는 두 아이템 i와 j사이의 유사도를 표현한다. " $\bullet$ "는 두 벡터의 내적을 나타내며, " $\|\|$ "는 벡터의 크기를 나타낸다. 코사인 계수는 두 대상간의 각도를 측정하여 유사한 정도를 판정하며, 값의 범위는 0에서 1사이의 값을 갖는다. 값이 1에 가까울수록 두 대상이 유사하다라고 해석된다.

유사도 산출의 세 번째 방법은 유클디언 거리계수(Euclidean Distance Coefficient)이다.

$$s_{a,u}=dist(a,u)=\sqrt{\sum_{i=1}^{n}(r_{a,i}-r_{u,i})^2}$$ 식4

식4는 벡터공간에서 두 사용간의 직선거리를 측정하여, 거리상 가까이 있는 유사한 사용자를 산출하는 수식이다. n은 전체 아이템의 수를 나타내며, 값의 범위는 0에서 무한대이다. 값이 0에 가까울수록 두 사용자 또는 두 아이템간의 유사도가 높다고 해석된다.

2.5 추천결과 예측

사용자 또는 아이템간의 유사도가 구해지면, 다음 단계로 유사도를 이용하여 추천을 적용하기 위한 예측기법을 적용한다. 예측기법은 유사도 측정 대상이 사용자 기반 협력적 추천 기법이냐 아이템 기반 협력적 추천 기법이냐에 따라 다소 차이를 보인다. 식5는 사용자 기반 협력적 추천에서의 예측 기법이다.

$$P_{a,i}=\overline{r_a}+\frac{\sum_{u=1}^{n}(r_{u,i}-\overline{r_u})*S_{a,u}}{\sum_{u=1}^{n}S_{a,u}}$$ 식5

식5에서 $\overline{P_{a,i}}$는 사용자 a에게 아이템 i를 추천하였을 때의 예측값을 나타내고, n은 가장 유사한 이웃의 크기를 나타내며, $\overline{r_{a,i}}$는 사용자 a가 모든 아이템들에 대하여 부여한 Rating의 평균을 의미한다. 사용자 기반 협력적 추천의 예측 방법은 추천 고객이 모든 아이템들에 대하여 Rating한 값의 평균을 Rating하지 않은 나머지 아이템들의 기본 선호도로 사용하고, 추천하고자 하는 아이템에 대해 유사한 이웃 사용자들이 부여한 선호도 여부에 유사도를 가중치로 부여하여 가산하는 방식이다. 다시 말해, 추천 고객의 평균 선호도에 유사 사용자들의 추천 아이템에 대한 선호도에 가중치를 반영한 값이라 할 수 있으며, 아이템들에 대해 예측을 수행한 후 수치 값을 내림차순으로 정렬하여 수치가 높은 상위 N개의 아이템을 고객에게 추천한다.

아이템 기반 협력적 추천에서의 예측 방법은 식6과 같다.

$$P_{u,i}=\frac{\sum_{all\ similar\ items,N}(S_{i,N}\times R_{u,N})}{\sum_{all\ similar\ items,N}(|S_{i,N}|)}=\frac{\sum_{N:\ items}(S_{i,N}\times R_{u,N})}{\sum_{N:\ items}(|S_{i,N}|)}$$ 식6

식6에서, $P_{u,i}$는 사용자 u 에게 아이템 i를 추천하였을 때의 예측값을 나타내며, S_{iN}은 i번째 아이템과 사용자가 구매한 아이템 N과의 유사도를 말한다. $R_{u,N}$은 사용자 u가 아이템 N에 제공한 Rating 정보를 나타낸다.

아이템 기반 협력적 추천 기법에서 예측 방법은 고객에게 추천될 아이템과 유사한 아이템들에 대해 부여된 선호도에 유사도를 가중치로 적용하는 방식이며, 추천 대상 아이템이 이웃 아이템과 유사도가 높고, 이웃 아이템에 부여된 Rating이 높을수록 추천 대상으로 선정될 확률이 높아진다.

3. 연구내용

3.1 연구목적 및 방법

기존의 추천 시스템은 추천을 적용하는 방법에 있어, 도메인의 특성을 고려하지 못하여 추천의 정확도를 떨어뜨리는 문제를 발생시켜왔다. 이러한 문제를 해결하고, 추천 시스템의 효율적인 적용을 위한 방안으로 본 논문에서는 추천 적용에 앞서 추천에 적합한 도메인 및 고객군을 평가할 수 있는 척도를 제안하고, 이를 통하여 추천을 위한 마케팅 도메인 및 고객군을 선정하고, 선행평가값을 높여 추천의 정확도를 높일 수 있는 방안을 제시한다. 구체적인 내용은 다음과 같다.

첫째, 추천 적용을 위한 선행평가 척도를 제안하고 선행 평가 척도간의 상관관계를 분석한다. 선행 평가척도란 앞서 간략히 언급한 바와 같이 추천을 적용하기 전에 과연 선정된 도메인 및 고객군이 추천이 잘 적용될 것인지 또는 잘 적용되지 않을 것인지 여부를 판정해 주는 평가 기준이라고 정의

할 수 있다. 본 연구에서 제안하는 선행 평가기준은 다음과 같으며, 세 가지 선행평가척도간 상관관계를 분석한다.

- 평균 사용자 유사도(Average User Similarity)
- 평균 아이템 유사도(Average Item Similarity)
- 밀집도(Density)

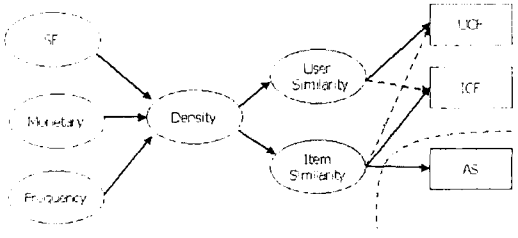
평균 사용자 유사도란 사용자 기반 협력적 추천 기법에서 사용자간의 유사도의 합을 사용자 쌍의 수로 나눈 평균값으로 정의되며, 평균 아이템 유사도는 아이템 기반 협력적 추천 기법에서 아이템간의 유사도의 합을 아이템 쌍의 수로 나눈 평균값으로 정의된다. 마지막으로 밀집도란 사용자가 서로 다른 아이템을 얼마나 많이 구매하였나를 의미하며, 각각의 사용자가 구매한 아이템의 종류의 합을 사용자와 아이템 공간의 크기로 나눈 것으로 정의된다.

둘째, 선정된 마케팅 도메인 및 고객군 내에서 추천의 정확도를 높이기 위한 방안으로 선행평가척도인 평균 유사도(Average Similarity) 또는 밀집도(Density)를 높이기 위한 방법을 제시하며, 그 내용은 다음과 같다.

- 도메인(Domain) 한정
- Support Filter(SF) 적용
- 고객 스코어(Score) 적용
  - Monetary, Frequency 등
- 기타 계층화 구조 활용 등

유사도 또는 밀집도를 높이기 위한 방법으로 임계값을 적용하여 Support Filter(SF: 구매 횟수), Monetary(구매 금액), Frequency(구매 빈도) 등의 임계값보다 낮은 수치를 갖는 사용자 또는 아이템을 제거하는 기법을 사용한다. 이러한 기법은 사용자와 아이템간 벡터 공간의 밀집도가 높아지고, 밀집도가 높아짐에 따라 동시 선호한 아이템 항목이 증가되어 평균 유사도가 높아진다는 가정에서 출발한다. 도메인의 한정은 전체 도메인에 비하여 사용자와 아이템 공간이 축소되어 유사도 및 밀집도에 영향을 줄 것이며, Monetary와 Frequency에 기준한 우량 고객에 대한 유사도 추출 및 추천 적용은 구매횟수, 구매금액 및 구매빈도의 임계치를 높임에 따라 고객이 구매한 아이템 종류가 많아져서 유사도와 밀집도에 영향을 미치게 된다. 계층화 구조의 활용은 아이템을 상위 개념으로 묶어 줌으로써 동일한 효과를 발생한다.

3.2 선행평가척도



[그림 3] 선행평가 척도 간의 상관관계

[그림 3]은 본 논문에서 제안하는 선행평가 척도 간의 상관관계를 나타낸다. SF, Monetary, Frequency가 Density에서 영향을 미치고, Density는 User Similarity나 Item Similarity에 영향을 미쳐 UCF(User-based Collaborative Filtering), ICF(Item-based Collaborative Filtering), AS(Association Rule)에 영향을 미칠 것이라는 내용이다. 상기 그림에서 우측 하단 점선 안에 있는 AS는 본 논문의 연구범위에 해당하지 않는 향후 연구로 선정하였으며, 이 이외의 영역에 대하여 실험을 수행하여 평가한다.

본 연구에서는 [그림 3]의 내용을 여섯 개의 가설로 설정하여 이를 실험을 통하여 평가한다. 각 가설에 대한 상세한 설명은 다음과 같다.

◆가설1. User Similarity에 의한 CF도메인 및 고객군 선정
- 가설: 사용자간 평균 Similarity가 높을수록 각 추천기법의 Precision이 높을 것이다.
- 가설수식

IF $Average\_Similarity_U(U_1,I_1) < Average\_Similarity_U(U_2,I_2)$
THEN
$Average\_Precision_{UCF}(U_1,I_1) < Average\_Precision_{UCF}(U_2,I_2)$
AND
$Average\_Precision_{ICF}(U_1,I_1) < Average\_Precision_{ICF}(U_2,I_2)$

가설 1은 User Similarity가 높으면, 해당 도메인 및 고객군에서 사용자 기반 협력적 추천(UCF)과 아이템 기반 협력적 추천(ICF)의 Precision이 높을 것이라는 의미이다. 가설수식에서 Average User Similarity는 다음과 같은 수식을 이용하여 산출된다.

$$Average\_Similarity_U(U,I) = \frac{\sum_{i=1}^{|U|}\sum_{i\neq j}^{|U|} Similarity(V(u_i,I),V(u_j,I))}{|U|^2-|U|} \quad \text{식7}$$

식7에서 Average_Similarity $_U$(U,I)은 아이템집합 I에 대해서 $u_i \in U$들의 벡터간 Similarity의 평균을 나타내며, Similarity($V(u_m,I),V(u_n,I)$)는 Vector $V(u_m,I),V(u_n,I)$사이의 Similarity를 나타내고, V(u,I)는 u가 아이템집합 I에 대해 Rating한 Vector를 의미한다.

◆가설2. Item Similarity에 의한 CF도메인 및 고객군 선정
- 가설: 아이템간 평균 Similarity가 높을수록 각각 추천기법의 Precision이 높을 것이다.
- 가설수식

IF $Average\_Similarity_I(U_1,I_1) < Average\_Similarity_I(U_2,I_2)$
THEN
$verage\_Precision_{UCF}(U_1,I_1) < Average\_Precision_{UCF}(U_2,I_2$
AND
$Average\_Precision_{ICF}(U_1,I_1) < Average\_Precision_{ICF}(U_2,I_2)$

가설 2는 Average Item Similarity가 높으면, 해당 도메인 및 고객군 안에서 사용자 기반 협력적 추천 및 아이템 기반 협력적 추천의 Precision이 높다는 것이다. 가설수식에서 Average User Similarity는 다음과 같은 수식을 이용하여 산출한다.

$$Average\_Similarity_f(U,I) = \frac{\sum_{u_m \in U}\sum_{i_m \in I} Similarity(V(u_m,I),V(u_n,I))}{|U|\cdot|U|} \quad \text{식8}$$

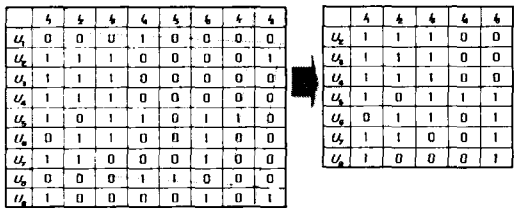
식8에서 $Average_Similarity_f(U,I)$은 고객집합 U에 대해서 $i_m \in I$들의 벡터간 Similarity의 평균을 나타내며, $Similarity(V(U,i_m),V(U,i_n))$는 Vector $V(U,i_m)$,$V(U,i_n)$사이의 Similarity를 나타내고, $V(U,i)$는 i가 고객집합 U에 의해 Rating된 Vector를 의미한다.

◆가설3. 밀집도(Density)에 의한 CF도메인 및 고객군 선정
◦ 가설: Density가 높으면 각 추천기법의 Precision이 높을 것이다.
◦ 가설수식
IF $Density(U_1,I_1) < Density(U_2,I_2)$
THEN
$Average\_Precision_{UCF}(U_1,I_1) < Average\_Precision_{UCF}(U_2,I_2)$
AND
$Average\_Precision_{ICF}(U_1,I_1) < Average\_Precision_{ICF}(U_2,I_2)$

가설 3은 밀집도(Density)가 높으면, 해당 도메인 및 고객군에서 사용자 기반 협력적 추천 및 아이템 기반 협력적 추천의 Precision이 높을 것이라는 것이다. $Density(U_1,I_1)$는 U_1와 I_1공간에서 데이터 밀집도를 나타낸다. 밀집도란 사용자와 아이템 공간에서 구매한 아이템 종류의 분포도를 나타내며, 밀집도가 낮을수록 데이터의 희박성(Sparsity)이 높아진다.

◆가설4. Support Filter가 밀집도에 미치는 영향
◦ 가설: 적절한 User수가 확보된다는 가정 하에, Support Filter가 높을수록 각 추천기법의 Precision이 높을 것이다.

가설 4는 Support Filter(SF)의 임계값을 높일수록 선호도가 많이 표시된 사용자와 아이템 공간만이 남게 되어 밀집도가 향상되고, 그 결과 사용자 기반 협력적 추천 및 아이템 기반 협력적 추천 기법의 Precision이 높아진다는 것이다. [그림 4]는 본 가설을 설명하고 있으며, SF 적용에 따른 차원이 축소되고 밀집해지는 현상을 보여주고 있다.



[그림 4] Support Filter(SF)에 의한 차원축소

◆가설 5,6 : Monetary, Frequency가 밀집도에 미치는 영향
◆가설 5. 우량고객 선정에 있어서 기준이 되는 Monetary를 높일수록 밀집도에 영향을 주어 각 추천기법의 Precision이 좋아질 것이다.
◆가설 6. 우량고객 선정에 있어서 기준이 되는 Frequency를 높일수록 밀집도에 영향을 주어 각 추천기법의 Precision이 좋아질 것이다.

가설 5와 가설 6은 우량고객 선정 기준인 Monetary와 Frequency를 높일수록 사용자가 구매하는 아이템의 종류의 다양해져서 해당 도메인의 밀집도와 유사도가 높아져 결국 사용자 기반 협력적 추천 및 아이템 기반 협력적 추천의 정확도가 높아질 것이라는 연구내용이다.

지금까지 제안된 선행평가 척도에 대하여 기술하였으며, 4장에서는 앞서 언급된 가설들을 실험을 통하여 평가하고, 선행평가 척도간의 상관관계 및 선행 평가 척도와 추천결과와의 상관관계 등을 분석한다.

4. 실험결과 및 분석

4.1 실험 데이터

실험 데이터 집합으로 백화점 오프라인 데이터 집합을 사용하여 실험을 수행하였다. 백화점 데이터 집합은 고객 11,300명과 상품 12,504개로 구성되어 있다. 훈련집합(Training Set) 대 실험집합(Test Set) 비율은 3 : 1로 선정하였으며, 훈련집합을 통하여 학습을 수행하고 실험집합을 예측 및 추천의 정확도를 측정하는데 사용하였다. 유사한 고객 및 아이템을 찾는 방법으로는 코사인 계수(Cosine Coefficient)를 사용하였다.

협력적 추천 기법은 일반적으로 각 고객이 부여한 선호도 정보를 이용하여 각 고객에게 개인화된 아이템을 추천한다. 본 논문에서 사용된 백화점 데이터 집합은 명시적 선호도 정보가 없으므로, 암시적 정보인 구매 정보를 활용하여 선호도 정보를 추출하였다. 각각의 아이템 항목에 대한 구매회수의 평균을 이용하여 평균 이상 구매한 아이템에 대하여 1.0, 평균 이하 구매한 아이템에 대하여 0.5, 비구매한 아이템에 대하여 0값을 할당하여 선호도 정보로 이용하여 추천기법을 적용하였다.

4.2 평가방법

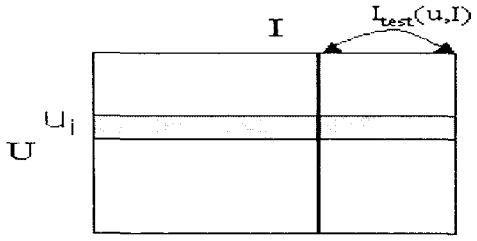
추천의 성능을 측정하는 방법으로는 Precision, Recall, 및 F(Sarwar et al, 2000) 측정법을 사용하였으며, 수식은 다음과 같다.

$$Precision = \frac{size\ of\ hit\ set}{size\ of\ topN\ set} = \frac{|test \cap topN|}{|N|} \quad \text{식9}$$

$$Recall = \frac{size\ of\ hit\ set}{size\ of\ test\ set} = \frac{|test \cap topN|}{|test|} \quad \text{식10}$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad \text{식11}$$

Precision(정확도)은 사용자에게 아이템이 N개 추천되었을 때 추천된 아이템 중에서 사용자가 좋아하는 아이템의 비율이고, Recall(재현율)은 사용자가 좋아하는 아이템 중 사용자가 좋아하는 추천된 아이템의 비율이다. F 측정법은 정확도와 재현율을 조합하여 하나의 평가치로 정의한 것으로, F값이 1에 근접해지면 정확도와 재현율의 값이 모두 좋아짐을 나타내며, 0에 근접해지면 정확도와 재현율 중 한 쪽이 나빠짐을 나타낸다.



[그림 5] UCF에서 Precision

[그림 5]는 Precision 측정 방법을 도해한 것으로, 고객(U)과 아이템(I) 공간에서 아이템 공간을 각 사용자별로 랜덤하게 3 대 1의 비율로 훈련집합과 테스트집합으로 나누어 훈련집합 중 사용자가 구매하지 않은 아이템을 추천하고, 테스트집합에 있는 아이템과 일치여부를 검사하여 일치하는 아이템을 히트집합(hit set)으로 사용하여 Precision을 계산한다.

본 논문에서는 먼저 각각 사용자에 대하여 Precision, Recall, F를 측정하고 나중에 각 측정값의 평균을 계산하여 추천 도메인의 측정치로 사용하였다. 사용자 기반 협력적 추천(UCF)에서 Averager User Precision은 식12와 같다.

$$Average\_Precision_{UCF}(U,I) = \frac{\sum_{u \in U} \frac{|I_{test}(u,I) \cap I_{top N}(u,I)|}{|I_{top N}(u,I)|}}{|U|} \quad \text{식12}$$

식 12에서 $Average_Precision_{UCF}(U,I)$은 각각의 사용자 $u \in U$에 대하여 아이템 I중 u가 Rating 하지 않은 아이템을 UCF로 추천하였을 때 Precision의 평균을 의미하고, $I_{topN}(u,I)$은 아이템집합 I중 u에게 추천된 상위 N개 아이템집합을 나타내며, $I_{test}(u,I)$은 아이템집합 I중 u의 test 아이템집합을 의미한다.

4.3 실험결과 및 분석

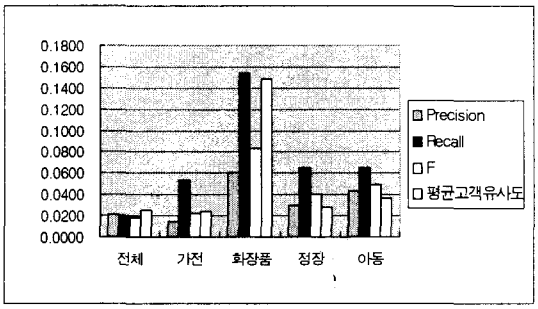
4.3.1 Average User Similarity에 의한 추천 도메인 및 고객군 선정

<표 1> 도메인별 User Similarity 분포도

유사도 \ 도메인	전체	가전	화장품	정장	아동
0	3516	4462	2563	4462	4105
0.0~0.1	973	15	78	13	136
0.1~0.2	403	199	579	161	306
0.2~0.3	39	153	624	135	229
0.3~0.4	18	72	516	88	111
0.4~0.5	1	23	300	48	37
0.5~0.6	0	13	158	26	22
0.6~0.7	0	6	83	13	4
0.7~0.8	0	4	38	2	0
0.8~0.9	0	1	8	1	0
0.9~1.0	0	1	3	1	0
1	0	1	0	0	0
총합계	4950	4950	4950	4950	4950
평균고객유사도	0.0245	0.0242	0.1482	0.0273	0.0364

<표 1>은 도메인별 두 고객 쌍간의 User Similarity 분포를 나타내고 있다. <표 1>에서 화장품 도메인이 다른 도메인에 비하여 유사도의 분포가 높은 수치까지 고루 분포되어 있고, 평균 고객유사도가 높은 것을 알 수 있다. 따라서, 본 가설 1에 의하여, 화장품 도메인 및 고객군 안에서 각 추천기법의 Precision이 높을 것으로 판단된다. [그림 6]은 고객에게 상위 5개의 아이템을 추천하였을 때 User-based CF의 추천 성능을 나타낸다.

측정값 \ 도메인	전체	가전	화장품	정장	아동
Precision	0.0215	0.0143	0.0608	0.0293	0.0429
Recall	0.0202	0.0536	0.1540	0.0650	0.0646
F	0.0186	0.0221	0.0835	0.0401	0.0492
평균고객유사도	0.0245	0.0242	0.1482	0.0273	0.0364



[그림 6] Top-5 추천에서 UCF 추천 성능

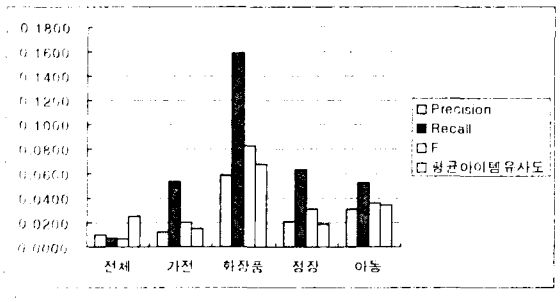
<표 2>가설1 분석(1)

측정값 \ 순위	Average User Similarity	UCF Precision
1	화장품	화장품
2	아동	아동
3	정장	정장
4	전체	전체
5	가전	가전

<표 2>는 Average User Similarity(AUS)와 UCF Precision의 수치를 각각 내림차순으로 정렬하여 요약한 것으로, AUS가 높은 순위와 동일하게 UCF 추천의 Precision 수치도 나타남을 볼 수 있다. 따라서, Average User Similarity가 높으면 UCF의 추천 성능 Precision이 높을 것이라는 가설 1을 만족시켜 주고 있다.

[그림 7]은 Item-based CF로 추천하였을 때 추천 성능을 나타내고 있으며, 실험을 분석하여 정리하면 <표 3>와 같다.

측정값 \ 도메인	전체	가전	화장품	정장	아동
Precision	0.0103	0.0128	0.0580	0.0211	0.0306
Recall	0.0078	0.0532	0.1583	0.0632	0.0526
F	0.0070	0.0203	0.0825	0.0311	0.0362
평균아이템유사도	0.0246	0.0147	0.0678	0.0184	0.0342



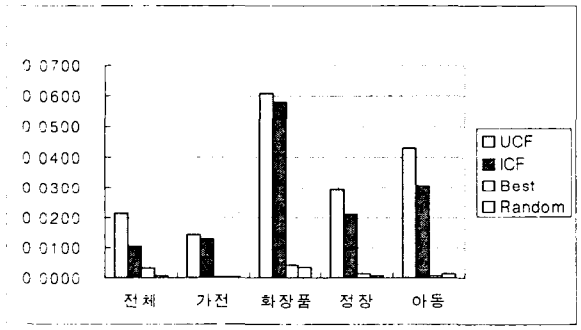
[그림 7] Top-5 추천에서 ICF 추천 성능

<표 3>가설1 분석(2)

측정값 / 순위	Average User Similarity	ICF Precision
1	화장품	화장품
2	아동	아동
3	정장	정장
4	전체	가전
5	가전	전체

<표 3>은 AUS와 ICF Precision 값을 내림차순으로 정렬하여 요약한 것이다. 표를 통하여 순위 1위에서 순위 3위까지 해당하는 화장품, 아동, 정장 도메인에서 Average User Similarity의 순위와 ICF의 Precision의 순위가 동일함을 알 수 있다. 따라서, 모든 도메인을 고려한다면 Average User Similarity의 순위와 ICF의 Precision의 순위가 일반적으로 유사하다고 평가할 수 있으므로, Average User Similarity가 높으면 ICF 추천성능의 Precision이 높다라는 가설 1을 일반적으로 만족시켜 주고 있다.

SF / 추천	전체	가전	화장품	정장	아동
UCF	0.0215	0.0143	0.0608	0.0293	0.0429
ICF	0.0103	0.0128	0.0580	0.0211	0.0306
Best	0.0032	0.0005	0.0042	0.0013	0.0007
Random	0.0008	0.0003	0.0035	0.0006	0.0013



[그림 8] Best, Random 추천과 CF 성능 비교

[그림 8]에서 Best 추천은 각 사용자별로 랜덤하게 추출된 훈련집합에서 베스트 목록을 산정하여, 각 사용자별로 사용자가 보지 않은 아이템에 대하여 베스트 목록에서 상위 N개를 추출하여 추천하는 방법을 사용하였으며, Random 추천은 랜덤하게 추출된 훈련집합에서 사용자가 보지 않은 아이템에 대하여 무작위로 N개를 추출하여 추천하는 방법을 사용하였다. [그림 8]의 추가 실험결과를 통해, CF가 Best나 Random추천에 비해 비교적 성능이 낮다는 것을 알 수 있으며, 차후 뒤따르는 가설 검증 실험을 통하여 Precision을 높일 수 있는 방법을 평가한다.

4.3.2 Average Item Similarity에 의한 추천 도메인 및 고객군 선정

<표 4>는 도메인별 두 아이템 쌍간의 Item Similarity 분포를 나타낸 것으로, <표 1>의 User Similarity에 비해 값의 분포가 폭 넓게 펼쳐져 있음을 알 수 있다. 그러나, 상위값을 갖는 분포대의 아이템 쌍의 구매내역을 분석해보면 대부분이 1개의 공통 구매한 아이템이 존재하는 경우가 대부분이어서 의미를 갖지를 못하여 실질적으로 비슷한 성격의 아이템이 존재하지 않아 ICF 추천의 성능이 떨어질 것임을 의미한다. 따라서, 본 도메인에서 보다 효율적 추천을 적용하기 위한 방안으로 Average Item Similarity를 이용하여 가장 추천이 잘 적용될 도메인 및 고객군을 선정한다.

<표 4> 도메인별 Item Similarity 분포도

도메인 / 유사도	전체	가전	화장품	정장	아동
0	642827	14387	1027	10256	20020
0.0~0.1	187	16	90	13	62
0.1~0.2	1683	81	152	80	288
0.2~0.3	3114	96	112	105	305
0.3~0.4	2855	63	58	57	277
0.4~0.5	3535	62	25	70	206
0.5~0.6	2923	52	13	63	149
0.6~0.7	963	11	5	17	72
0.7~0.8	3347	40	3	28	159
0.8~0.9	1480	22	0	16	99
0.9~1.0	44	0	0	0	5
1	6945	48	0	26	94
총합계	669903	14878	1485	10731	21736
평균아이템유사도	0.0246	0.0147	0.0678	0.0184	0.0342

<표 5> 가설2 분석

측정값 / 순위	Average ItemSimilarity	ICF Precision	UCF Precision	Density
1	화장품	화장품	화장품	화장품
2	아동	아동	아동	아동
3	전체	정장	정장	정장
4	정장	가전	전체	가전
5	가전	전체	가전	전체

<표 5>는 Average Item Similarity(AIS)와 ICF 및 UCF 추천기법의 Precision 수치를 내림차순으로 정렬하여 요약한 것으로, AIS가 높으면 일반적으로 ICF나 UCF의 추천기법의 Precision이 높음을 나타내고 있다. ICF 추천기법의 Precision 패턴은 실험 도메인에서 1개의 공통 구매하는 아이템이 유독 많이 존재하여 AIS가 높게 나타난 '전체'란 도

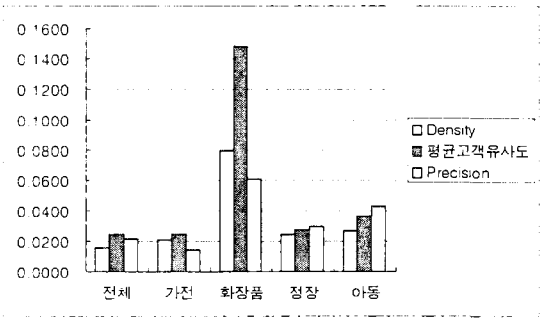
메인을 제외하면 화장품, 아동, 정장, 가전 순위가 되어 패턴이 AIS와 비슷하게 된다. 또한 UCF 추천기법의 Precision은 5개 도메인 중 3개 도메인 화장품, 아동, 가전에서 비슷한 패턴을 보여주고 있어 Average Item Similarity가 높음에 따라 ICF 및 UCF 추천 기법의 Precision이 높을 것이라는 가설을 뒷받침해 주고 있다.

그러나, ICF 도메인 선정에서 AIS뿐만 아니라 Density를 고려한 것이 더욱 정확한 가설이 된다고 평가된다. AIS와 Density가 동시에 높은 도메인을 순위화면 조건에 해당되지 않는 '전체' 도메인은 제외되고, 화장품, 아동, 정장, 가전의 순위가 형성되어 UCF 및 ICF 추천의 Precision 순위와 동일해지기 때문이다.

4.3.3 Density에 의한 추천 도메인 및 고객군 선정

백화점 데이터로 밀집도에 따른 UCF 및 ICF의 성능을 실험하였다. [그림 9]는 평균고객유사도(Average User Similarity)와 Density와의 관계를 나타낸다.

측정값 \ 도메인	전체	가전	화장품	정장	아동
Density	0.0155	0.0209	0.0791	0.0247	0.0263
평균고객유사도	0.0245	0.0242	0.1482	0.0273	0.0364
Precision	0.0215	0.0143	0.0608	0.0293	0.0429



[그림 9] Density, 평균고객유사도, Precision (UCF)

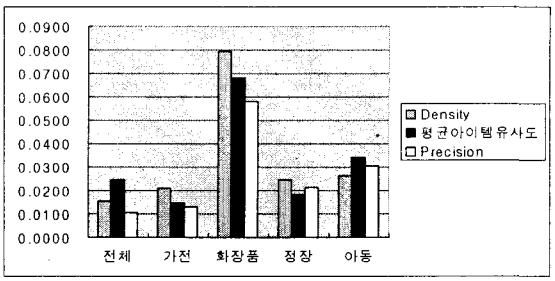
<표 6> 가설3 분석(1)

순위 \ 측정값	Density	Average User Similarity	UCF Precision
1	화장품	화장품	화장품
2	아동	아동	아동
3	정장	정장	정장
4	가전	전체	전체
5	전체	가전	가전

<표 6>은 Density, Average User Similarity 및 UCF 추천의 Precision 수치를 내림차순으로 정렬한 것으로서, 일반적으로 Density가 높으면 UCF 추천기법의 Precision이 높을 것이라는 가설을 뒷받침한다.

[그림 10]은 평균 아이템 유사도(Average Item Similarity)와 Density를 비교한 것이며, <표 7>은 실험을 정리하여 요약한 도표이다.

Density	0.0155	0.0209	0.0791	0.0247	0.0263
평균아이템유사도	0.0246	0.0147	0.0678	0.0184	0.0342
Precision	0.0103	0.0128	0.0580	0.0211	0.0306



[그림 10] Density, 평균아이템유사도, Precision (ICF)

<표 7> 가설3 분석(2)

순위 \ 측정값	Density	Average Item Similarity	ICF Precision
1	화장품	화장품	화장품
2	아동	아동	아동
3	정장	정장	정장
4	가전	전체	가전
5	전체	가전	전체

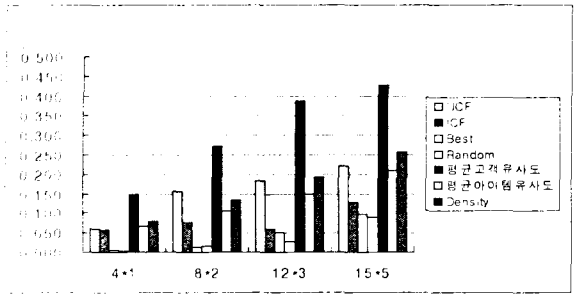
<표 7>은 Density와 Average Item Similarity의 수치를 내림차순으로 정렬한 것으로, Density가 높음에 따라 ICF 추천 기법의 Precision이 높을 것이라는 가설을 뒷받침해 주고 있다.

따라서, UCF 및 ICF 추천 기법 모두에서 밀집도가 증가하면 각 추천기법의 Precision이 증가할 것이라는 가설 3이 입증되었다고 할 수 있다.

4.3.4 Support Filter가 선행평가 척도 및 추천에 미치는 영향

Support Filter(SF)를 적용하기 위한 도메인으로는 앞서 실험에서 비교적 높은Precision을 나타낸 화장품 도메인으로 선정하여 실험을 하였다. [그림 11]은 SF에 따른 UCF 및 ICF 추천기법의 Precision을 실험한 결과이며, 추가실험으로 Best, Random추천 기법의 성능을 분석하여 보여 주고 있다. SF에서 '*' 문자의 앞에 있는 숫자는 User 필터 임계값을 나타내며, '*' 문자 뒤의 숫자는 Item 필터 임계값을 나타낸다. 실험결과를 통해, SF가 증가함에 따라 선행평가 척도인 Average User Similarity, Average Item Similarity, Density가 증가하여 UCF 추천 기법의 Precision은 증가하고, ICF 추천 기법의 Precision은 일반적으로 증가함을 알 수가 있다. 따라서, 이는 SF가 증가함에 따라 각각 추천기법의 Precision이 높아질 것이라는 가설 4를 밝혀 주고 있다. 또한 실험을 통하여 CF 추천 기법이 Best나 Random 추천 기법에 비해서 성능이 우수함을 알 수 있었다.

추천 \ SF	4*1	8*2	12*3	15*5
UCF	0.061	0.156	0.183	0.221
ICF	0.058	0.076	0.060	0.126
Best	0.004	0.012	0.051	0.097
Random	0.003	0.013	0.025	0.090
평균고객유사도	0.1482	0.2724	0.3870	0.4286
평균아이템유사도	0.0678	0.1051	0.1486	0.2089
Density	0.0791	0.1336	0.1919	0.2577

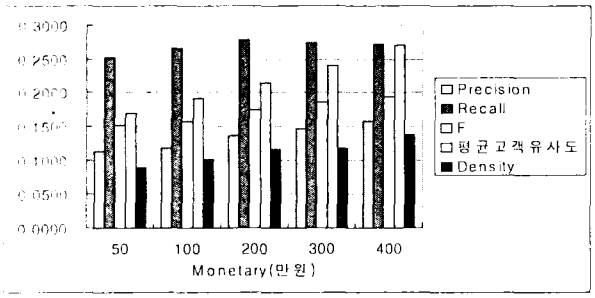


[그림 11] SF에 따른 선행평가 척도 및 추천 Precision

4.3.5 Monetary가 선행평가 척도 및 추천에 미치는 영향

Monetary를 적용하기 도메인으로 역시 4.3.4절과 같이 화장품 도메인을 사용하였다. 화장품 도메인은 각각 고객별 평균 구매금액이 144만원에 해당하며, Monetary의 임계값으로는 50만원부터 400만원까지 적용하여 실험을 하였다.

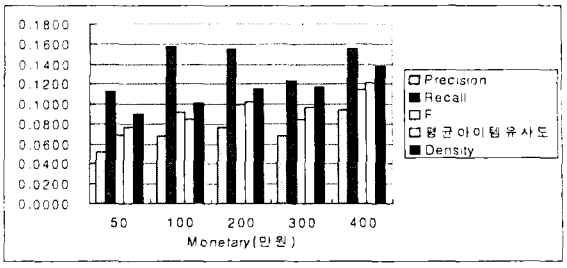
측정값 \ Monetary	Precision	Recall	F	평균고객유사도	Density	고객수
50	0.1133	0.2519	0.1514	0.1685	0.0893	639
100	0.1172	0.2661	0.1568	0.1908	0.1007	385
200	0.1363	0.2782	0.1743	0.2138	0.1155	196
300	0.1462	0.2735	0.1859	0.2400	0.1166	111
400	0.1574	0.2710	0.1938	0.2694	0.1377	64



[그림 12] Monetary에 따른 UCF 추천 성능

[그림 12]는 Monetary를 높임에 따른 UCF 추천 기법의 성능을 실험한 결과를 나타내고 있으며, Monetary를 높임에 따라 선행평가 척도인 Density와 Average User Similarity가 증가하고 Precision이 높아짐을 알 수 있다. 이는 가설 5를 뒷받침해 준다.

측정값 \ Monetary	Precision	Recall	F	평균아이템유사도	Density	고객수
50	0.0520	0.1125	0.0688	0.0761	0.0893	639
100	0.0680	0.1570	0.0916	0.0839	0.1007	385
200	0.0760	0.1548	0.0989	0.1016	0.1155	196
300	0.0680	0.1227	0.0840	0.0969	0.1166	111
400	0.0938	0.1557	0.1142	0.1214	0.1377	64



[그림 13] Monetary에 따른 ICF 추천 성능

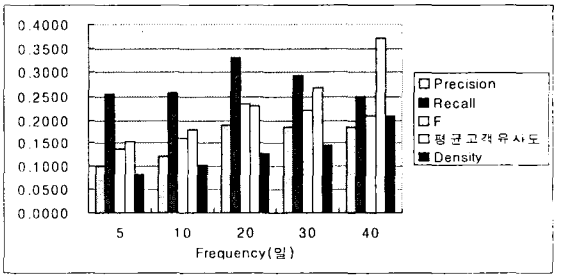
[그림 13]은 Monetary가 증가함에 따라 ICF 추천 기법의 성능을 보여 주고 있으며, Monetary가 증가함에 따라 Density가 증가하고, Density의 증가가 일반적으로 Average Item Similarity를 증가시켜 일반적으로 ICF의 Precision이 증가하고 있음을 나타내고 있다.

4.3.6 Frequency가 선행평가 척도 및 추천에 미치는 영향

실험 도메인으로 화장품 도메인을 사용하였으며, 평균 구매 방문 횟수는 12일(회)이며, Frequency 임계값은 5회부터 40회까지 사용하였다.

[그림 14]는 Frequency의 증가에 따른 UCF 추천 성능 및 관련 수치를 보여 주고 있다. 실험 결과를 분석하면, Frequency가 증가함에 따라 선행평가 척도인 Density와 Average User Similarity가 증가하고, UCF 추천 기법의 Precision이 일반적으로 증가하고 있으며, 이는 가설 6을 뒷받침해 주고 있다.

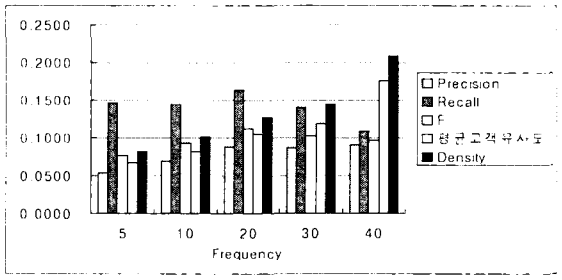
측정값 \ Monetary	Precision	Recall	F	평균아이템유사도	Density	고객수
50	0.0520	0.1125	0.0688	0.0761	0.0893	639
100	0.0680	0.1570	0.0916	0.0839	0.1007	385
200	0.0760	0.1548	0.0989	0.1016	0.1155	196
300	0.0680	0.1227	0.0840	0.0969	0.1166	111
400	0.0938	0.1557	0.1142	0.1214	0.1377	64



[그림 14] Frequency에 따른 UCF 추천 성능

[그림 15]는 Frequency에 따른 ICF 추천 기법의 추천 성능 및 관련 수치를 보여주고 있다. [그림 15]의 실험결과를 분석해 보면, Frequency를 증가시킴에 따라 선행평가 척도인 Density와 Averager Item Similarity가 증가하고 ICF 추천기법의 Precision이 일반적으로 증가하고 있음을 알 수 있다. 따라서, Frequency가 증가함에 따라 선행평가 수치가 증가하여 UCF 및 ICF 추천 기법의 Precision이 일반적으로 증가한다고 말할 수 있다.

측정값 / Monetary	Precision	Recall	F	평균고객유사도	Density	고객수
5	0.0540	0.1467	0.0765	0.0671	0.0818	804
10	0.0700	0.1442	0.0921	0.0815	0.1007	440
20	0.0880	0.1628	0.1122	0.1044	0.1263	176
30	0.0861	0.1403	0.1028	0.1178	0.1440	72
40	0.0897	0.1080	0.0966	0.1755	0.2084	29



[그림 15] Frequency에 따른 ICF 추천 성능

제 5 장 결론 및 향후연구

본 논문에서는 추천을 적용하기에 앞서 마케팅 도메인 및 고객군 선정을 위한 선행 평가척도로서 평균 사용자 유사도(Average User Similarity), 평균 아이템 유사도(Average Item Similarity), 밀집도(Density)를 제안하였고, 선행평가 척도값을 높여 추천의 정확도가 높은 도메인을 선정할 수 있는 방안으로 Support Filter, Monetary, Frequency를 제안하였다. 실험결과를 통하여 Support Filter, Monetary, Frequency가 밀집도를 높여 주고, 밀집도가 높을 경우 Average User Similarity 및 Average Item Similarity가 높아짐을 알 수 있었으며, 이에 따라 고객기반 CF 및 아이템기반 CF 추천 기법의 Precision 성능이 일반적으로 증가함을 알 수 있었다. 따라서, 실험 도메인에서 추천 선행평가 척도로서 Density, Average User Similarity, Average Item Similarity의 적용이 가능하다고 평가되어지며, Support Filter, Monetary, Frequency를 이용하여 선행평가 척도값을 높여 추천의 Precision이 향상된 마케팅 도메인 및 고객군 선정이 가능하다고 평가되어진다. 또한 제안한 선행평가 척도 Average Similarity는 각 사용자별 K-NN을 찾는 단계까지 계산하지 않음으로써, 추천 마케팅 도메인 및 고객군 선정 수행 시간이 단축된다는 이점을 갖는다.

향후연구로서, 4장에서 실험평가를 위해 사용한 평가방법의 정확성을 높이기 위해 다양한 도메인으로의 확장이 필요하다. 또한, 선행 평가 척도로서 선호도 정보뿐만 아니라, 데모그래픽 정보를 활용하는 방안 및 연관규칙에 대해 본 연구의 적용 방안 등도 향후 연구 과제이다. 향후연구를 통해 가설이 보다 객관적으로 밝혀지고, 여러 추천기법에서 적용이 가능해진다면, 추천 선행평가 척도로서 효용적 방안이 될 것이다.

참고문헌

Badrul Sarwar, George Karypis, Joseph Konstan, John Riedel, "Item-Based Collaborative Filtering Recommendation Algorithms", GroupLens Research Group/Army HPC Research Center, 2001.

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

Badrul Sarwar, Joshep Konstan, Al Borchers, Jon Herlocker, Brad Miller, John Riedl, "Using Filtering Agents to Improve Prediction Quality in the GroupLens Research Collaborative Filtering System", In Proceedings of CSCW '98, Seattle, WA.

Marko Balabanovic, Yoav Shoham, "Content-Based, Collaborative Recommendation", 1977.

Pazzani Michael, "A Framework for Collaborative, Content-Based and Demographic Filtering", 1999.

Paul Resnick, Neophytos Iacovou,Mitesh Suchak, Peter Bergstrom, John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In preoceedings of CSCW '94 Chapel hill, NC.

Upendra Shardanand, Pattie Maes, "Social Information Filtering: Algorithms for Automating Word of Mouth", In Proceedings of the CHI'95.

Weiyang Lin, Sergio A.Alvarez, Carolina Ruiz, "Collaborative Recommendation via Adaptive Association Rule Mining", 1999.