

협업 필터링 기반 상품 추천에서의 평가 횟수와 성능*

이흥주** · †김종우*** · 박성주**

Number of Ratings and Performance in Collaborative Filtering-based Product Recommendation*

Hong Joo Lee** · Jong Woo Kim*** · Sung Joo Park**

■ Abstract ■

The Collaborative Filtering (CF) is one of the popular techniques for personalization in e-commerce storefronts. For CF-based recommendation, every customer needs to provide subjective evaluation ratings for some products based on his/her preference. Also, if an e-commerce site recommends a new product, some customers should rate it. However, there is no in-depth investigation on the impacts on recommendation performance of two number of ratings, i.e. the number of ratings of an individual customer and the number of ratings of an item, even though these are important factors to determine performance of CF methods.

In this study, using publicly available EachMovie data set, we empirically investigate the relationships between the two number of ratings and the performance of CF. For the purpose, three analyses were executed. The first and second analyses were performed to investigate the relationship between the number of ratings of a particular customer and the recommendation performance of CF. In the third analysis, we investigate the relationship between the number of ratings on a particular item and the recommendation performance of CF. From these experiments, we can find that there are thresholds in terms of the number of ratings below which the recommendation performances increase monotonically. That is, the number of ratings of a customer and the number of ratings on an item are critical to the recommendation performance of CF when the number of ratings is less than the thresholds, but the value of the ratings decreases after the numbers of ratings pass the thresholds. The results of the experiments provide insight to making operational decisions concerning collaborative filtering in practice.

Keyword : Collaborative Filtering, Product Recommendation, Personalization, e-Commerce

논문접수일 : 2005년 8월 25일 논문게재확정일 : 2006년 3월 14일

* 이 논문은 2005년 한양대학교 일반연구비 지원으로 연구되었음(HY-2005-G).

** 한국과학기술원, 테크노경영대학원

*** 한양대학교 경영학부

† 교신저자

1. 서론

전자상거래에서 고객에게 적합한 혹은 관심 있어 할만한 한 상품을 개인화하여 제공하는 것은 고객의 상품 검색 노력을 줄여 줄 뿐만이 아니라, 고객의 충성도 제고에도 도움을 주기 때문에 고객관계관리 측면에서도 중요하게 인식되고 있다[1, 12, 13, 19]. 고객의 선호도를 파악하여 고객에게 개인화된 서비스의 제공과 상품을 추천하기 위해 많은 연구들과 추천 기법, 추천 시스템들이 제시되어 왔다[1-3, 6-8, 10, 12, 14-21, 23, 28, 29]. 현재 상용화된 추천 시스템들에서 가장 많이 활용되고 있는 추천 기법은 협업 필터링(Collaborative Filtering, CF)이며[4, 5, 11, 15, 22], 해당 고객과 선호도가 유사한 고객들의 선호도를 활용하여 추천할 상품을 선정하는 기법이다[6, 25, 26, 29].

협업 필터링에서는 고객들이 제시한 상품들에 대한 평가 점수에 의존하여 고객간의 유사도와 상품에 대한 예상 평가 점수를 계산하여 이를 바탕으로 상품을 추천한다. 따라서 고객의 상품 평가 점수로 이루어진 고객-상품 행렬이 상품 추천을 위한 기본적인 데이터가 된다. 이러한 고객-상품 행렬은, 고객이 모든 상품에 대하여 평가 점수를 제시하는 것이 불가능하기 때문에, 희소 행렬(sparse matrix) 형태를 가지게 된다. 이러한 고객-상품 행렬의 희소성을 극복하기 위한 방안으로 사용자나 상품을 클러스터링하거나 구매 데이터를 활용한 기법들을 활용하여 선호도 데이터가 매우 희소한 상황에서 추천 성능을 개선시키는 방안들이 연구되고 있다[13, 23]. 또한 이러한 고객-상품 행렬의 희소성은 *cold start* 문제를 야기한다. 협업 필터링의 *cold start* 문제란 상품에 대한 평가 점수가 적은 고객에 대해서는, 그리고 고객들의 평가가 적은 상품에 대해서는 좋은 추천 성능을 제공하지 못하는 문제점을 의미한다[24, 26]. 이러한 문제를 해결하기 위해서 고객의 선호도 표시 데이터가 적은 초기에는 고객의 인구통계학적 정보에 기반을 둔 추천 방안과 베스트셀러와 같은 유명한 상품을 추천하는 방안을

활용하는 추천 전략에 대한 연구가 이루어졌다[24].

협업 필터링의 추천을 위해서 활용되는 기본 데이터가 고객-상품 행렬 형태로 저장된 평가 점수들이므로, 고객의 평가 횟수가 추천 성능에 직접적으로 영향을 주는 것을 직관적으로 추측할 수 있다. 하지만, 이러한 평가 횟수와 추천 성능에 대한 심층적이고 실증적인 연구는 아직까지 부족한 형편이다. 본 연구에서는 협업 필터링의 연구에서 많이 활용되는 EachMovie 데이터 집합을 이용하여 고객의 평가 횟수와 추천 성능간의 관계에 대하여 실험을 수행하였다. 첫 번째로 평가 횟수가 다른 고객 집단간의 추천 성능의 차이에 대한 실험을 수행하였다. 두 번째 실험은 신규 고객이 평가를 거듭하면서 추천 성능이 어떻게 향상되는지 에 대한 실험을 수행하였다. 마지막으로 몇 번 이상 선호도가 표시된 상품을 고객에게 추천하는 것이 협업 필터링의 추천 성능을 높일 수 있는가에 대한 실험을 수행하였다. 본 논문의 구성은 다음과 같다. 2장에서는 협업 필터링 기법과 관련 연구들을 살펴본다. 3장에서는 평가 횟수와 추천 성능에 대한 세 가지 실험 설계와 실험결과를 소개한다. 4장에서 실험의 의의와 본 연구의 결론을 제시한다.

2. 관련 문헌

2.1 협업 필터링의 기본 구조

협업 필터링의 첫 번째 단계는 고객의 선호도 데이터를 가지고 고객-상품 행렬을 구성하는 것이다. 전통적인 협업 필터링 알고리즘에서의 입력 데이터는 사용자 n 명의 상품 m 개에 대한 선호도 평가 점수나 구매이력 자료이다. 이 고객-상품간의 $n \times m$ 행렬을 S 라고 하면, S_{ik} 는 고객 i 의 상품 k 에 대한 평가점수이고, \bar{S}_i 는 고객 i 의 평가점수 평균이다.

협업 필터링의 두 번째 단계는 고객간의 유사도를 구하는 것이다. 고객간의 유사도를 피어슨 상관계수 형태로 구하는 계산하는 식은 식 (1)과 같다. 식 (1)은 고객 i, j 간의 상관계수 r_{ij} 를 구하는 식으

로, 상관계수 r_{ij} 는 두 고객의 선호도가 유사한 경우에는 1에 가까운 값을 가지게 되고, 상반된 선호도를 갖는 경우에는 -1에 가까운 값을 가지게 된다.

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

식 (2)는 고객-상품간의 행렬을 고객별로 벡터로 표현하여, 두 고객의 유사도를 두 고객 벡터간의 코사인 값으로 계산하는 식이다. 이 경우에는 고객 a , b 를 m 차원의 벡터로 표현하게 된다[17].

$$Similarity(\vec{a}, \vec{b}) = \cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|} \quad (2)$$

식 (1)이나 식 (2)를 통해 계산된 고객들 간의 유사도를 통해 유사 고객(neighbor)을 선정하게 된다. 협업 필터링의 마지막 단계는 상품에 대한 고객의 선호도 점수를 예측하는 것이다. 상품에 대한 고객의 선호도 점수 예측은 다음의 식 (3)를 통해서 이루어진다. 식 (3)은 고객 i 의 상품 k 에 대한 선호도 점수인 P_{ik} 를 예측하는 식으로, $Rater(k)$ 는 상품 k 를 평가한 고객의 집합을 의미한다.

$$P_{ik} = \bar{S}_i + \frac{\sum_{l \in Rater(k)} (S_{lk} - \bar{S}_l) r_{il}}{\sum_{l \in Rater(k)} |r_{il}|} \quad (3)$$

2.2 협업 필터링 관련 연구

협업 필터링에 대한 연구 중의 많은 연구들이, 새로운 협업 필터링 방안을 제시하거나 기존의 방안을 수정한 기법들을 제시하고, 이들을 기존 방안들과 성능 비교를 하고 있다. Breese et al.[6]과 Gonzalez-Caro et al.[9]는 다양한 협업 필터링 알고리즘의 성능을 비교하였으며, Mild and Natter[19]는 협업 필터링과 회귀모델의 성능을 비교하였다. 사용자 기반의 협업 필터링에서 가장 문제시 되고 있는 부분은 사용되는 데이터의 희소성(sparcity)과 협업 필터링 알고리즘의 확장성(scalability)

이기 때문에[15, 17, 25, 26, 29], 이와 관련된 연구들이 많이 진행되었다. 협업 필터링에 적용되는 알고리즘은 고객-상품 행렬 S 에서 고객과 상품의 차원이 증가함에 따라 더욱 많은 컴퓨팅 역량을 필요로 하기 때문에 추천 성과를 감소시키지 않으면서 사용되는 상품과 고객의 차원을 줄이기 위한 연구들이 진행되었다. Latent Semantic Indexing(LSI) 기법을 활용하여 고객-상품 행렬(S)의 차원을 줄이는 방안들이 제시되었으며[17, 24], 이 방안은 $n \times m$ 고객-상품 행렬(S)이 singular value decomposition에 의해서 랭크 k 의 근사 행렬로 대체한 후에 협업 필터링을 수행하는 것이다. Linden et al.[17]은 고객 차원의 축소를 위해 사용자의 무작위 추출, 평가회수가 적은 사용자의 제거, 유사 사용자들의 클러스터링 같은 방법을 활용하였으며, 상품 차원의 축소를 위해 평가회수가 많은 상품이나 평가 횟수가 적은 상품을 제거하는 방법을 활용하였다. Sarwar et al.[24]는 사용자와 상품의 차원이 추천 성과에 미치는 영향에 대해서 조사하였으며, 효율적인 인덱싱 기법의 활용을 통해 추천 알고리즘의 확장성을 향상시킬 수 있음을 보였다.

또한, 협업 필터링에서 사용자에게 대한 예측치를 계산할 때 활용되는 네이버(neighbor)들의 규모가 협업 필터링 알고리즘의 계산 효율성과 추천 성과에 영향을 미친다. 추천 성과를 저하시키지 않는 적절한 네이버들의 규모를 계산하였으며, MovieLens 데이터 집합에서는 80명에서 120명 정도의 네이버가 적절한 것으로 파악되었으며 EachMovie 데이터 집합에서는 170명에서 220명 정도의 네이버 규모가 적절한 것으로 파악되었다[24, 25]. 이와 함께, 추천 방안의 성과를 측정하기위해 학습 집합과 테스트 집합으로 데이터를 나누는 적절한 비율에 대한 연구도 수행되었으며 전체집합에서 0.8의 데이터를 학습 집합으로 나누는 것이 적절한 것으로 파악되었다[24, 25].

사용자의 평가 횟수가 추천 성능에 밀접하게 관련되어 있음에도 불구하고, 이들 간의 관계에 대한 연구는 상당히 부족한 형편이다. 실제로, 평가 횟수

〈표 1〉 협업 필터링에서의 계수 선택

	데이터 집합 사용자 수, 상품 수	고객의 최소 선호도 표시 수	상품의 최소 피 선호도 표시 수	네이버 (Neighbor) 의 수	추천 상품 수
Mild & Natter (2002)	EachMovie 61,007명, 419개	3개	50회	10~80명 (최적)	-
Sarwar et al. (2000)	MovieLens 943명, 1682개	20개	1회	80~120명 (최적)	10개
Sarwar et al. (2001)	MovieLens 943명, 1682개	20개	1회	30명 (최적)	-
Ansari et al. (2000)	EachMovie 2000명, 340개	1개	1회	-	-
Breese et al. (1998)	EachMovie 4119명, 1623개	-	-	-	-

가 추천 성능에 결정적인 역할을 함에도 불구하고, 협업 필터링의 성능 비교를 수행한 많은 연구들에서조차도 원래의 데이터 집합에서 실험에 활용할 데이터 집합을 추출하면서, 몇 번 이상의 선호도를 표시한 사용자의 데이터를 활용하였는지, 몇 번 이상 선호도가 표시된 상품을 활용하였는지에 대하여 아무런 근거를 제시하지 않거나, 어떤 연구들에서는 이들에 대한 언급이 빠진 경우도 많이 있다. <표 1>은 관련 연구들에서의 실험에 사용한 데이터 집합과 실험에 포함된 고객의 최소 선호도 표시 수, 상품의 최소 선호도 표시 수, 유사고객의 수, 추천 상품의 수 등을 요약한 것이다. <표 1>에서 볼 수 있듯이 대부분의 연구들이 자의적으로 계수를 선택하여 데이터를 선정하였으며, 어떠한 기준으로 실험용 데이터 집합을 선정하였는지에 언급하지 않고 있다. 또한 <표 1>에서 ‘-’로 표시된 경우는, 해당 데이터를 제시하지 않은 경우이다.

협업 필터링에서는 고객의 상품에 대한 선호도 표시 데이터가 적은 경우에는 추천 성능이 좋지 않은 *cold start* 문제가 발생한다[24, 26]. 이 문제를 해결하기 위해서 전자상거래 사이트나 추천 시스템

에서는 사용자의 선호도 표시 데이터가 적은 초기에는 고객의 인구통계학적 정보에 기반을 둔 방안과 베스트셀러와 같은 유명한 상품을 추천하는 기법들이 활용하는 추천 전략을 통해 추천을 수행한다[24]. 그 후에 고객의 선호도 표시 데이터가 많아지게 되면 협업 필터링 방안으로 추천 전략을 변경하여 추천을 수행한다. 그러나 이러한 해결책에도 불구하고, 언제 협업 필터링 방안으로 추천 전략을 전환할 것인지에 대한 물음이 남게 된다. *Cold start* 문제에 관한 많은 연구들이 초기 추천 전략을 제시하는데 중점을 두고 있어, 고객의 선호도 표시 데이터의 수가 초기에서 점점 증가하는 경우에 협업 필터링으로 언제 전환할지에 대한 연구에는 많은 관심을 두지 않았다.

본 연구에서는 다루고자 하는 평가 횟수와 추천 성능에 대한 관계에 대하여, 구체적인 연구 질문들은 다음과 같다.

- 고객의 평가 횟수에 따라 협업 필터링의 추천 성능이 어떻게 달라지는가?
- 고객의 평가 횟수가 증가함에 따라 협업 필터

링의 추천 성능은 어떻게 향상되는가?

- 협업 필터링의 안정적인 추천 성과를 얻기 위해서는 고객이 몇 회 이상 선호도를 표시한 상품을 추천하여야 하는가?

3. 평가 횟수와 추천 성능 간의 관계

본 연구에서는 협업 필터링에서의 평가 횟수와 추천 성능 간의 연관관계를 파악하기 위하여, 협업 필터링 연구에서 많이 사용되는 EachMovie 데이터 집합을 활용하여 실험을 수행하였다[6, 19]. EachMovie 데이터 집합은 72,916명의 고객이 1,628개의 영화에 대해 총 2,811,983회의 선호도를 표시한 데이터 집합이며, 선호도 표시는 0점부터 5점까지 6단계로 이루어졌다. 다음의 절들에서 수행된 실험에는 <표 2>에 표기된 것 같은 데이터 집합이 활용되었다.

협업 필터링의 성능 측정방안으로는 기존 연구에서 많이 활용되는 척도인 예측치와 실제 선호도간의 차이에 기반을 둔 Mean Absolute Error(MAE)와 Root Mean Squared Error(RMSE)를 활용하였으며, 두 지표는 다음과 같은 식을 통해 계산된다.

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (4)$$

$$SE = \sqrt{\frac{\sum_{i=1}^N (p_i - q_i)^2}{N}} \quad (5)$$

N 은 예측하여야하는 상품의 총 수이며, p_i 는 i 상품에 대해 고객이 입력한 실제 선호도 값이며 q_i 는 i 상품에 대해 협업 필터링에 의해 계산된 선호

도 예측치이다.

다음 3개의 절에서 평가 횟수와 추천 성능간의 관계를 파악하기 위해 수행된 3개의 실험을 자세하게 설명하도록 한다. 3.1절은 고객의 선호도 표시 수와 추천 성능간의 관계를 파악하기 위한 실험이며, 3.2절은 고객의 초기 시점부터 선호도 표시 수 증가에 따른 추천 성능의 변화를 파악하기 위한 실험이다. 3.3절은 상품에 대한 선호도 표시 수와 추천 성능간의 관계를 파악하기 위해 수행된 실험이다.

3.1 고객의 선호도 표시 수와 협업 필터링 성능간의 관계에 관한 실험

3.1.1 데이터 집합과 실험 방안

실험의 수행을 위해 <표 2>에 표시된 것처럼 EachMovie 데이터 집합에서 4개 이상의 영화에 선호도를 표시한 고객들 중 무작위로 추출된 10%의 고객을 실험에 활용하였다. EachMovie 데이터 집합에서 무작위로 선정된 데이터 집합은 모두 5,748명의 고객이 249,840 개의 선호도를 표시한 집합이며, 고객의 선호도 표시 횟수의 평균은 43.37회, 표준편차는 51.70이다. 실험을 위해 선정된 데이터 집합들을 고객의 선호도 표시 횟수를 기준으로 <표 3>과 같이 각각 10개의 그룹으로 구분하였다.

<표 3>에 표기된 데이터의 희소수준은

$$1 - \frac{Nonzero\ Entries}{Total\ Entries} \quad (6)$$

로 정의되며[26], 여기서 $Total\ Entries$ 는 (고객의 수 \times 상품의 수)이며 $Nonzero\ Entries$ 는 고객들의 총 선호도 표시 횟수이다.

<표 2> 실험 데이터 집합

데이터 집합	실험	고객 선정 기준	고객 수	선호도 표시수	평균 표시수	표준편차
EachMovie	-	-	72,916	2,811,983	-	-
사용자 10%	3.1	4개이상 투표	5,748	249,840	43.47	51.70
사용자 2.5%	3.2	4개이상 투표	1,437	66,634	46.37	50.64
사용자 5%	3.3	-	2,874	134,197	46.69	53.74

<표 3> 데이터 집합의 그룹화

그룹	선호도 표시 횟수 범위	선호도 표시 횟수 평균	편차	고객수	총 선호도 표시 횟수	회소수준
1	4~6	4.8547	0.7662	606	2942	0.9970
2	7~9	7.9822	0.8199	505	4031	0.9951
3	10~14	11.8134	1.4691	643	7596	0.9927
4	15~20	17.3996	1.7331	588	10231	0.9893
5	21~26	23.5579	1.6763	509	11991	0.9855
6	27~34	30.2385	2.3126	524	15845	0.9814
7	35~46	40.0608	3.5981	608	24357	0.9755
8	47~63	54.4125	4.8393	572	31124	0.9666
9	64~94	77.5317	9.1004	568	44038	0.9524
10	95~817	156.3456	77.1566	625	97716	0.9039

각 그룹에 속한 고객의 선호도 표시 데이터의 70%를 무작위로 선정하여 고객간의 유사도 계산에 활용하며, 나머지 30%의 데이터에 대해 선호도를 예측하였다. 실험은 다음과 같은 단계를 거쳐 수행되었다.

① 학습 집합과 테스트 집합 선정

<표 3>에 표시된 각 그룹에서 무작위로 선정된 70%의 고객 선호도 데이터를 고객간의 유사도와 상품에 대한 예측 선호도를 계산하기 위한 학습 집합으로 선정하고, 나머지 30%의 고객 선호도 데이터는 추천 성과를 측정하기 위한 테스트 집합으로 활용하기 위하여 구분하였다.

② 고객간의 유사도 계산

①에서 선정된 학습 집합과 고객간의 상관관계를 측정하는 식 (1)을 활용하여 고객간의 유사도를 계산하였다.

③ 상품에 대한 선호도 예측치 계산

②에서 계산된 고객간의 유사도와 식 (3)을 활용하여, ①에서 선정된 테스트 집합에 속해있는 평가 점수에 대한 예측 선호도를 계산하였다.

④ 협업 필터링의 추천 성능 측정

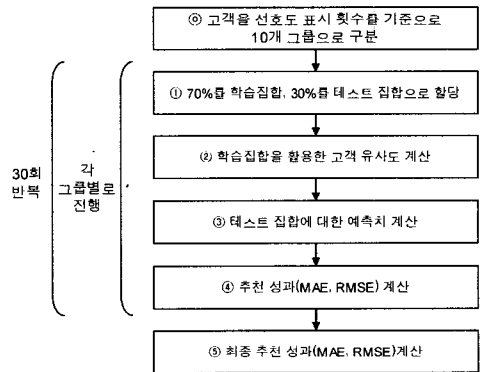
③에서 계산된 테스트 집합에 속한 예측 선호도와 실제 사용자가 입력한 선호도를 이용하여, 식

(4), 식 (5)의 MAE와 RMSE를 계산하였다.

⑤ 반복수행

위의 과정을 30회 반복하여 수행하였다.

실험 절차에 대한 도식적인 표현은 [그림 1]과 같다.

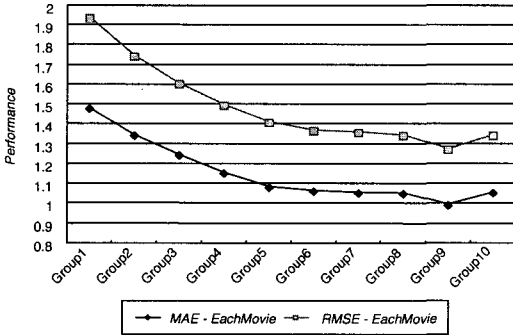


[그림 1] 고객의 선호도 표시 수와 성능간의 관계에 대한 실험 절차

3.1.2 실험 결과

위 실험의 결과를 EachMovie 데이터 집합의 각 그룹별 추천 성과로 표시한 것이 [그림 2]이다. [그림 2]에 표시된 MAE와 RMSE값은 30회 반복수행을 통해 나온 MAE, RMSE 결과의 평균치이다. MAE와 RMSE 값이 그룹 1에서 그룹 5까지 지속적으로 감소하는 것을 볼 수 있으며, 그룹 9를 제외

한 그룹 5부터 그룹 10까지의 MAE와 RMSE값은 크게 차이가 없는 것을 볼 수 있다.



[그림 2] 고객의 선호도 표시 수와 추천 성능 간의 관계

위 그룹들의 MAE 값의 평균과 표준편차를 정리한 것이 <표 4>이며, 그룹별 추천 성능의 차이를 알아보기 위하여 one-way ANOVA 테스트를 수행하였다. 이 테스트의 귀무가설은 다음과 같다.

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k$$

μ_i 는 i 번째 그룹의 MAE 평균값을 뜻하며, 귀무가설의 뜻은 그룹들의 MAE 평균값이 동일하다는 것이다. <표 4>에 정리된 것처럼 EachMovie 데이터에 대해 테스트한 결과 F값이 1807.506이고, p값이 0.000이어서 귀무가설은 기각되었다. 이는 10개 그룹의 추천 성과가 차이가 난다는 것으로 볼 수 있다. Duncan 테스트 결과에서 알 수 있듯이 그룹 1에서 그룹 6까지는 유의한 차이로 MAE 값이 모두 단조 감소하며, 그룹 6, 7, 8, 10은 유사한 값을 나타내고 그룹 9가 가장 낮은 MAE 값을 갖는다고 볼 수 있다.

그룹의 사용자 선호도 표시수가 적을수록 추천 성과는 안 좋은 것으로 볼 수 있으며, 일정 수준 이상의 평가 점수를 확보하면 추천 성능의 개선이 거의 없는 것을 볼 수 있다. EachMovie 데이터 집합에서는 고객의 선호도 표시수가 34개를 넘으면 추천 성능이 안정적인 값을 나타내었다.

<표 4> 고객 평가 횟수와 추천 성능간의 관계

	MAE		F-value	Duncan 테스트
	평균	표준편차		
그룹 1	1.4823	0.0406	1807.506 (p-value =0.000)	그룹 1 > 그룹 2 > 그룹 3 > 그룹 4 > 그룹 5 > 그룹 6, 7, 8, 10 > 그룹 9
그룹 2	1.3449	0.0308		
그룹 3	1.2458	0.023		
그룹 4	1.1533	0.0187		
그룹 5	1.0863	0.0149		
그룹 6	1.0652	0.0129		
그룹 7	1.0557	0.0105		
그룹 8	1.0525	0.0073		
그룹 9	0.9936	0.0064		
그룹10	1.0545	0.0032		

따라서 협업 필터링을 활용하여 추천을 수행하는 전자상거래 사이트에서는 고객들로부터 많은 수의 선호도 표시 데이터를 확보하기 위해 많은 비용과 자원을 투입하는 것 보다는 일정 수준의 선호도 표시 데이터만을 확보하여 사용하는 것이 상품추천에서 효율적일 수 있다. 반대로 협업 필터링의 수행을 통해 개인화된 상품을 추천하기 위해서는 선호도 표시 데이터의 수가 적어도 일정 개수 이상 되어야 안정적인 추천 성과를 보장할 수 있다는 것을 의미한다.

3.2 고객의 선호도 표시 수에 따른 최적 초기 추천 시점에 관한 실험

3.2.1 데이터 집합과 실험 방안

고객의 선호도 표시 수에 따른 협업 필터링의 최적 초기 추천 시점에 관한 실험의 수행을 위해 EachMovie 전체자료 집합에서 무작위로 고객 2.5%를 추출한 데이터 집합을 활용하였다. EachMovie 집합에서 무작위로 선정된 데이터 집합은 1,437명의 고객이 66,634회의 선호도를 표시한 집합이며(희소수준 0.9715), 고객의 선호도 표시회수 평균은 46.37회, 표준편차는 50.64이다.

EachMovie 데이터 집합을 활용한 실험은 다음과 같은 단계를 거쳐 수행되었다.

① 테스트 고객의 선정

선정된 실험집합의 고객 중 선호도 표시 회수가 110회 이상 130회 이하인 고객 50명을 선출하여 테스트 고객으로 선정하였으며, 이외의 다른 모든 고객 데이터는 학습 집합으로 활용하였다.

② 초기화 및 중단

고객의 평가 횟수를 나타내는 $i(1 \leq i \leq 100)$ 를 1로 초기화하였다. 이 초기화 단계가 30회 이상 반복된 후에 실험을 중단하였다.

③ 학습 집합과 테스트 집합으로 분리

테스트 고객 집합의 고객 선호도 자료를 평가 일자 순서로 정렬하여 가장 최근 10개의 선호도 표시 데이터를 테스트 집합으로 선정하였다. 나머지 고객 선호도 자료에서 i 개 만큼의 선호도 자료를 무작위로 추출하여 ①단계에서 생성된 학습 집합에 첨가하였다.

④ 고객간의 유사도 계산

③단계에서 생성된 학습 집합을 가지고 식 (1)을 활용하여 고객간의 유사도를 계산하였다.

⑤ 상품에 대한 예측 선호도 계산

③단계에서 생성된 테스트 집합에 대해 ④단계에서 생성된 고객 유사도와 학습 집합을 가지고 식 (3)을 활용하여 고객의 예측 선호도를 계산하였다.

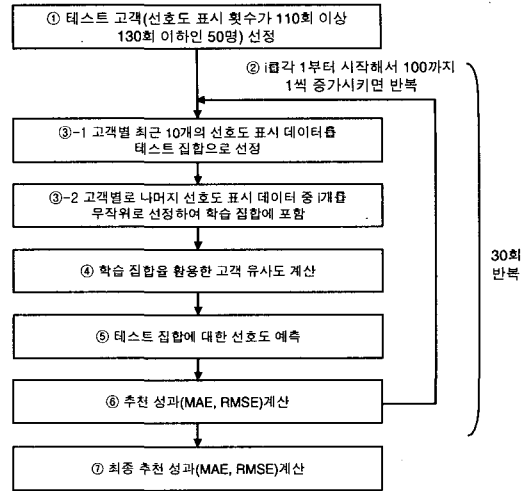
⑥ 협업 필터링의 성능 측정

③단계에서 생성된 테스트 집합에 속한 실제 사용자가 입력한 선호도와 ⑤단계에서 생성된 예측 선호도를 비교하여 MAE와 RMSE를 계산하였다.

⑦ 반복

i 의 값을 1씩 늘리고, ③단계로 넘어간다. 만약 1증가된 i 의 값이 100을 넘을 경우에는 ②단계로 이동하였다.

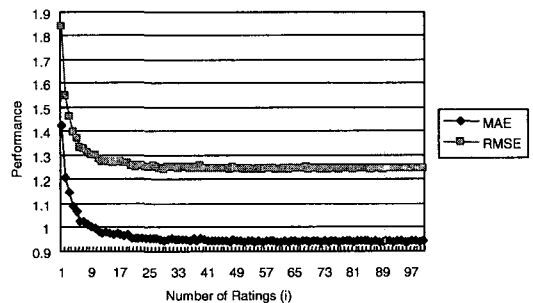
두 번째 실험 절차에 대한 도식적인 표현은 [그림 3]과 같다.



[그림 3] 고객의 선호도 표시 수에 따른 최적 초기 추천 시점에 관한 실험 절차

3.2.2 실험 결과

[그림 4]는 두 번째 실험 결과이며, i 값이 1에서 100까지 증가하는 경우의 추천 성과추이를 보여주고 있다. EachMovie 데이터 집합에서는 고객이 입력한 선호도를 1개에서 20개까지 활용하여 예측하였을 때에는 MAE값이 1.42에서 0.95까지 지속적으로 감소하였다.



[그림 4] 고객의 선호도 표시 수에 따른 최적 초기 추천 시점에 관한 실험

고객의 선호도 표시 수가 20~30개를 넘을 경우에는 추천 성능이 일정범위의 값을 갖도록 안정되며, 고객의 선호도 표시수가 30개 이상으로 증가하더라도 추천 성능이 개선되지 않는다는 것을 알 수 있다. 고객의 선호도 표시 수에 따른 최적 초기 추

천 시점에 관한 실험의 결과는 데이터의 특성에 따라 유동적이기는 하지만 사용자가 20~30개 이상의 상품에 선호도를 표시하였을 때가 그 이하만큼의 상품에 선호도를 표시하였을 때보다 추천 성과가 안정적인 것으로 볼 수 있다. 이는 3.1절의 실험결과와 유사하다고 볼 수 있다. 3.1의 실험에서는 고객의 선호도 표시 수가 21~26개인 그룹 5부터 추천 성과가 안정적이 되었다.

많은 전자상거래 사이트들이 고객의 선호도 데이터가 적은 시기에는 *cold start* 문제를 해결하기 위하여 인구통계학적 정보나 일반적으로 고객이 많이 구매하는 상품을 추천하는 방안을 활용하고 있다. 위의 실험결과를 활용하여, 고객에 대한 상품 추천 전략을 협업 필터링으로 전환하는 적절한 시점을 선택하는 기준으로 활용할 수 있다. 협업 필터링을 활용하여 상품을 추천하는 전자상거래 사이트에서는 자신의 데이터를 가지고 위와 유사한 실험의 수행하여, 협업 필터링에 의한 최적 초기 추천 시점을 정하는 것이 필요하다.

3.3 상품의 피 선호도 표시 수와 협업 필터링 성능 간의 관계에 관한 실험

3.3.1 데이터 집합과 실험 방안

실험의 수행을 위해 EachMovie 전체집합에서 무작위로 5%의 고객을 선택한 데이터 집합을 활용하였다. EachMovie 데이터 집합에서 무작위로 선정된 데이터 집합은 모두 2,874명의 고객이 134,197회의 선호도를 표시한 집합이며(희소수준 0.9713), 고객의 선호도 표시 횟수 평균은 46.69회, 표준편차는 53.74이다. 데이터 집합에서 영화에 대해 고객이 선호도를 표시한 최소 횟수는 1회이며, 최대 표시 횟수는 1,843회이다. 실험은 다음과 같은 단계를 거쳐 수행되었다.

① 학습 집합, 테스트 집합 선정

실험 집합에서 70%의 고객 선호도 데이터를 고객간의 유사도와 상품에 대한 예측 선호도 계산을 위한 학습 집합으로 선정하였고, 나머지 30%의 고

객 선호도 데이터를 테스트 집합으로 선정하였다.

② 고객간의 유사도 계산

①단계에서 학습 집합으로 선정된 데이터를 가지고 식 (1)을 활용하여 고객간의 유사도를 계산하였다.

③ 상품에 대한 예측 선호도 계산

①단계에서 학습 집합으로 선정된 데이터와 ②단계에서 계산된 고객간의 유사도를 가지고, 식 (3)을 활용하여 테스트 집합에 속한 평가 점수에 대한 예측 선호도를 계산하였다.

④ 협업 필터링의 성능 측정

①단계에서 생성된 테스트 집합에 속한 실제 고객이 입력한 선호도와 ③단계에서 생성된 예측 선호도를 비교하여 MAE와 RMSE를 계산하였다.

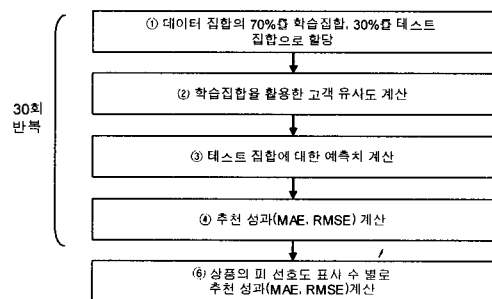
⑤ 반복

위 실험을 30회 반복 실시하였다.

⑥ 결과 분석

⑤단계까지의 실험이 종료된 이후에, 상품의 피 선호도 표시 수에 따라 MAE와 RMSE를 분석하였다.

[그림 5]는 세 번째 실험 절차에 대한 도식적인 표현이다.

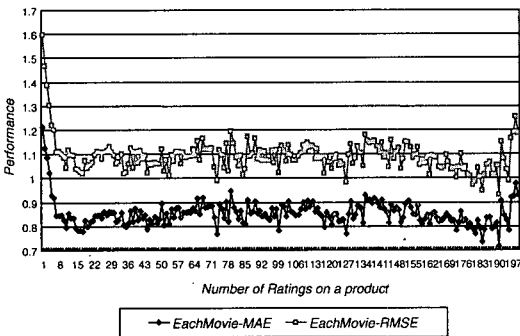


[그림 5] 상품의 피 선호도 표시 수와 성능간의 관계에 대한 실험 절차

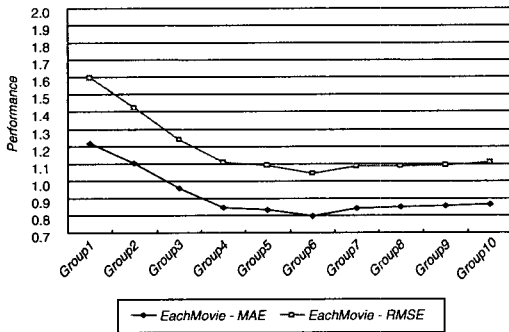
3.3.2 실험 결과

[그림 6](a)은 EachMovie 데이터 집합에서 상품의 피 선호도 표시 수가 1인 경우부터 200인 경우

까지의 MAE와 RMSE를 나타낸 것이다. 상품의 피 선호도 표시수가 1부터 10까지 증가하는 경우에 MAE와 RMSE가 급격히 떨어지는 것을 볼 수 있다. 상품의 피 선호도 표시수가 10을 넘어서면 MAE가 0.7과 0.9 수준에서 변동하지만 안정적인 성능을 보인다고 할 수 있다. [그림 6](a)에는 표시가 되지 않았지만 상품의 피 선호도 표시수가 200을 넘어 최대 표시 회수인 1843까지 증가하는 경우에도 10부터 200까지와 비슷한 결과를 보인다.



(a) 상품에 대한 선호도 표시 수



(b) 데이터 그룹

[그림 6] 상품에 대한 선호도 표시 수와 추천 성능 간의 관계

상품의 피 선호도 표시 수에 따른 추천 성능을 통계적으로 검증하기 위하여 <표 5>처럼 선호도 표시 수에 따라 데이터를 그룹화한 후에, 각 그룹별로 MAE 평균값의 차이를 검증하였다.

[그림 6](b)는 <표 5>에 따라 그룹화된 EachMovie

데이터 그룹들의 추천 성능을 나타낸 것이다. 그룹 1, 2, 3의 MAE 평균값이 다른 그룹들의 MAE 평균값보다 큰 것을 알 수 있으며, 다른 그룹들의 MAE 평균값은 유사한 것으로 볼 수 있다.

<표 5> 실험 집합의 그룹화

그룹	선호도 표시 횟수의 범위	상품의 수(30회 반복실험 총계)
1	1	63,597
2	2~3	77,259
3	4~6	70,730
4	7~9	51,622
5	10~14	56,676
6	15~21	52,648
7	22~36	54,358
8	37~74	52,201
9	75~200	53,047
10	201~1830	62,124

<표 6> 상품에 대한 투표수와 추천 성능간의 관계

MAE	EachMovie			
	Mean	std	F-value	Duncan Test
그룹 1	1.2143	0.1357	68.097 (p-value = 0.000)	Group 1 > Group 2 > Group 3 > Group 4, 5, 7, 8, 9, 10 > Group 6
그룹 2	1.1051	0.1203		
그룹 3	0.9569	0.0869		
그룹 4	0.8444	0.0821		
그룹 5	0.8293	0.0767		
그룹 6	0.7960	0.0835		
그룹 7	0.8389	0.0980		
그룹 8	0.8503	0.1325		
그룹 9	0.8539	0.1376		
그룹 10	0.8637	0.1242		

각 그룹들의 MAE의 평균과 표준편차를 정리한 것이 <표 6>이다. 그룹들의 추천 성능에 차이가 있는지를 one-way ANOVA로 테스트하였다. EachMovie 데이터 집합에 대한 귀무가설은 10개 그룹의 MAE 평균값이 같다는 것이었으며, F값이 68.097이고 p값이 0.000이어서 귀무가설은 기각되었다. Duncan 테

스트 결과에서 알 수 있듯이 그룹 1에서 그룹 4까지는 유의한 차이로 MAE 값이 모두 단조 감소하며, 그룹 4, 5, 7, 8, 9, 10은 유사한 값을 나타내고 그룹 6이 가장 낮은 MAE 값을 갖는다고 볼 수 있다. EachMovie 데이터의 경우에는 상품의 피 선호도 표시수가 15~21개를 넘었을 때 추천하여주는 것이 안정적인 추천 성능을 보이는 것으로 볼 수 있다.

그룹 내의 상품의 피 선호도 표시수가 적을수록 추천 성능이 안 좋은 것으로 볼 수 있으며, 상품에 대한 선호도 표시 수를 일정 수준 이상 확보하면 추천 성능의 감소는 크지 않은 것을 볼 수 있다. 전자상거래 업체의 입장에서는 신규상품의 경우에 실험에서 파악된 수준이하로 평가된 상품을 협업 필터링을 통해 추천하는 것은 적절하지 않다고 볼 수 있다. 상품을 추천하기 위해서는 실험에서 파악된 수준만큼의 선호도를 표시해주는 평가단을 운영하는 것도 추천에 도움을 줄 수 있다. 이 결과를 활용하여 고객들로부터 충분히 평가받지 못한 상품들을 제거하여 고객-상품 행렬의 차원을 축소시킬 수 있으며, 축소된 행렬로 협업 필터링을 수행하는 경우에는 협업 필터링의 확장성 문제를 어느 정도 경감시킬 수 있다.

4. 결 론

본 논문에서는 협업 필터링에 추천 성능과 고객의 선호도 평가 횟수와 상품의 피 선호도 표시 횟수의 관계에 대하여 연구를 수행하였다. 논문에서 수행한 세 가지의 실험을 통하여 고객의 선호도 평가 횟수와 상품의 피 선호도 표시 횟수가 적은 경우에는 평가 횟수가 증가하면, 추천 성능도 함께 향상되는 것을 볼 수 있었다. 하지만, 평가 횟수가 일정한 수준을 넘으면 추천 성능은 더 이상 향상되지 않는 것을 볼 수 있었다. 따라서 협업 필터링을 활용하는 전자상거래 업체가 안정적인 추천 성능을 확보하기 위해서는 고객 선호도 데이터를 적절한 최소 수준 이상으로 유지하여야 한다. 하지만, 일정한 수준 이상의 선호도 표시는 추천 성능에 추가적

인 이익을 가져오지 않으므로, 일정한 수준 이상의 추가적인 선호도 정보 획득을 위한 노력은 시간과 비용 측면에서 낭비가 된다.

본 연구 결과를 실제에 적용하는데 있어서 유의할 점은, 추천 성능에 향상의 한계가 되는 평가 횟수가 데이터 집합에 의존적이라는 것이다. 따라서 본 연구 결과를 실제 적용하기 위해서는 자신의 데이터 집합을 가지고 본 연구에서 실행한 것과 유사한 실험을 통해서 한계치를 확인하는 과정이 필요하다.

참 고 문 헌

- [1] 김재경, 서지혜, 안도현, 조운호, "협업 필터링 기법을 활용한 개인화된 상품 추천방법론 개발에 관한 연구", 「한국지능정보시스템학회논문지」, 제8권, 제2호(2002), pp.139-157.
- [2] 김종우, 이정미, "인터넷 상점에서 개인화 광고를 위한 장바구니 분석 기법의 활용", 「경영과학」, 제17권, 제3호(2000), pp.19-30.
- [3] Ansari, A., S. Essegai, and R. Kohli, "Internet Recommendation Systems," *Journal of Marketing Research*, Vol.37, No.3(2004), pp.363-375.
- [4] Autonomy, <http://www.autonomy.com>, 2004.
- [5] BeFree, <http://www.befree.com>, 2004.
- [6] Breese, J.S., D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Technical Report*, MSR-TR-98-12, Microsoft Research, 1998.
- [7] Chen, Y. and E. George, "A Bayesian model for Collaborative Filtering," Technical Report, Statistics Department, University of Texas at Austin, 2000.
- [8] Cho, Y.H., J.K. Kim, and S.H. Kim, "A Personalized Recommender System based on Web usage Mining and Decision Tree

- Induction," *Expert Systems with Applications*, Vol.23, No.3(2002), pp.329-242.
- [9] Gonzalez-Caro, C.N., M.L. Calderon-Benavides, J.J. Perez-Alcazar, J.C. Garcia-Diaz, and J. Delgado, "Towards a More Comprehensive Comparison of Collaborative Filtering Algorithms," In A.H.F. Laedner and A. L. Oliveira (eds.), *SPIRE 2002, Lecture Notes in Computer Science 2476*, (2002), pp.248-253.
- [10] Gupta, O., M. Digiovanni, H. Norita, and K. Goldberg, "Jester 2.0 : Evaluation of a new Linear time Collaborative Filtering Algorithm," *22nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, (1999), pp.291-292.
- [11] Gustos, <http://www.gustos.com>, 2004.
- [12] Kim, J.K., Y.H. Cho, W.J. Kim, J.R. Kim, J.R., and J.H. Suh, "A Personalized Recommendation Procedure for Internet Shopping Support," *Electronic Commerce Research and Applications*, Vol.1(2002), pp.301-313.
- [13] Kim, J.W., H.J. Lee, and S.J. Bae, "Data Sparsity and Collaborative Filtering-based E-Commerce Personalized Recommendation," *Working Paper*, Hanyang University, 2004.
- [14] Kim, J.W., B.H. Lee, M.J. Shaw, H. Chang, and M. Nelson, "Application of Decision Tree Induction Techniques to Personalized Advertisements on Internet Storefront," *International Journal of Electronic Commerce*, Vol.5, No.3 (2001), pp.45-62.
- [15] Konstan, J.A, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, and J. Riedl, "GroupLens : Applying Collaborative Filtering to Usenet News," *Communication of the ACM*, Vol.40, No.3(1997), pp.77-87.
- [16] Lee, C.H., Y.H. Kim, and P.K. Rhee, "Web Personalization Expert with Combining Collaborative Filtering and Association Rule Mining Techniques," *Expert Systems with Applications*, Vol.21, No.3(2001), pp.131-137.
- [17] Linden, G., B. Smith, and J. York, "Amazon.com Recommendations : Item-to-item Collaborative Filtering," *IEEE Internet Computing*, Vol.7, No.3(2003), pp.76-80.
- [18] Malone, T.W., K.R. Grant, F.A. Turbak, S.A. Brobst, and M.D. Cohen, "Intelligent Information Sharing Systems," *Communications of the ACM*, Vol.30, No.5(1987), pp.390-402.
- [19] Mild, A. and M. Natter, "Collaborative Filtering or Regression Models for Internet Recommendation Systems?," *Journal of Targeting, Measurement and Analysis of Marketing*, Vol.10, No.4(2002), pp. 304-313.
- [20] Mobasher, B., R. Cooley, and J. Srivastava, "Automatic Personalization based on web usage Mining," *Communication of the ACM*, Vol.43, No.8(2000), pp.142-151.
- [21] Mulvenna, M.D., S.S. Anand, and A.G. Buchner, "Personalization on the Net using Web Mining," *Communication of the ACM*, Vol.43, No. 8(2000), pp.122-125.
- [22] Net Perception, <http://www.netperceptions.com>, 2004.
- [23] Roh, T.H., K.J. Oh, and I. Han, "The Collaborative Filtering Recommendation based on SOM Cluster-indexing CBR," *Expert Systems with Applications*, Vol.25(2003), pp.413-423.
- [24] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommendation Algorithms for E-Commerce," *Proceedings of EC'00, Minneapolis*, (2000), pp.158-167.
- [25] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms," *Proceedings of WWW10*,

- Hong Kong, (2001), pp.285-295.
- [26] Schein, A.I., A. Popescul, L.H. Ungar, and D.M. Pennock, "Methods and Metrics for Cold-Start Recommendations," *Proceedings of the ACM SIGIR'02*, Tampere, Finland, (2002), pp.253-260.
- [27] Shardanand, U. and P. Maes, "Social Information Filtering : Algorithms for Automating 'Word of Mouth'," *Proceedings of Conference on Human Factors in Computer Systems*, (1995), pp.210-217.
- [28] Rashid, A.M., I. Albert, D. Cosley, S.K. Lam, S.M. McNee, J.A. Konstan, and J. Riedl, "Getting to Know You : Learning New User Preferences in Recommender Systems," *Proceedings of the ACM IUI'02*, San Francisco, (2002), pp.127-134.
- [29] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens : An Open Architecture for Collaborative Filtering of Netnews," *Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work*, New York, (1994), pp. 175-186.