Journal of The Korea Society of Computer and Information
Vol. 17 No. 4 , April 2012

2012-17-4-3-2

협업필터링에서 포괄적 성능평가 모델

유 석 종*

A Comprehensive Performance Evaluation in Collaborative Filtering

Seok Jong Yu *

요 약

대규모의 상품을 다루는 전자상거래 시스템에서 개인화된 추천은 필수적인 기능이 되고 있다. 대표적 추천 알고리즘인 협업필터링은 내용기반 추천에 비하여 뛰어난 추천성능을 제공해 주고 있으나, 희박성, 신규 아이템 문제(Cold-start), 확장성 등의 근본적인 한계를 갖고 있다. 본 연구에서는 추가적으로 협업필터링이 목표 대상자에 따라 비일관된 예측 능력의 차이를 보이는 추천 성능의 편차 문제를 제기하고자 한다. 추천성능의 편차는 기존의 Mean Absolute Error(MAE)에 의해서는 측정되기 어려우며 또한 정확도, 재현율 지표와도 독립적으로 평가되고 있다. 협업알고리즘의 정확한 성능평가를 위해서 본 연구에서는 MAE, MAE 편차, 정확도, 재현율을 포괄적으로 평가할 수 있는 확장 성능평가모델을 제안하고 이를 클러스터링 기반 협업필터링에 적용하여 성능을 비교 분석한다.

▸ Keyword : 추천시스템, 협업필터링, 성능평가, MAE

Abstract

In e-commerce systems that deal with a large number of items, the function of personalized recommendation is essential. Collaborative filtering that is a successful recommendation algorithm, suffers from the sparsity, cold-start, and scalability restrictions. Additionally, this work raises a new flaw of the algorithm, inconsistent performance of recommendation. This is also not measurable by the current MAE-based evaluation that does not consider the deviation of prediction error, and furthermore is performed independently of precision and recall measurement. To

• 제1저자 : 유석종 • 교신저자 : 유석종
• 투고일 : 2011. 12. 08, 심사일 : 2012. 01. 03, 게재확정일 : 2012. 02. 10.
* 숙명여자대학교 컴퓨터과학부(Dept. of Computer Science, Sookmyung Women's University)

evaluate the collaborative filtering comprehensively, this work proposes an extended evaluation model that includes the current criteria such as MAE, Precision, Recall, deviation, and applies it to cluster-based combined collaborative filtering.

▸ Keyword : Recommender System, Collaborative Filtering, Performance Evaluation, MAE Reduction

I. 서론

온라인 상에서 정보 공유, 통신, 전자상거래 등 기능과 활동이 활발해짐에 따라 정보의 양이 급격히 증가하고 있으며 키워드 기반 정보 검색의 한계가 발생하고 있다. 이를 보완하기 위하여 고객이 필요한 정보를 능동적으로 제시해주는 개인화된 추천 기능의 역할이 중요해지고 있으며, 특히 Amazon.com과 같이 음악, 영화, 책 등 상품의 수가 지속적으로 증가하는 대규모 전자상거래 사이트에서는 필수적이다. 개인화된 추천 방법에는 고객 프로필에 근거하는 내용기반 추천(content-based recommendation)과 많이 구매되는 상품을 추천하는 통계적 방법(statistical recommendation), 그리고 협업필터링(collaborative filtering)이 있다[1-5]. 내용기반 추천은 사용자 프로필에 추천에 필요한 충분한 자료가 포함되어 있어야 하는 특성이 있고 상품과 사용자의 수가 증가할수록 확장성의 문제가 발생한다[6]. 평가기록에 근거하는 협업필터링은 90년대 중반부터 활발히 연구되어 오고 있는 사회형 추천 알고리즘으로 고객과 구매 성향이 유사한 집단을 찾아 선호할만한 미경험 아이템을 추천해준다. 협업필터링의 추천의 질은 유사 선호도를 갖는 이웃 집단에 따라 결정되며, 이웃 탐색 방법으로 kNN(k-Nearest Neighbor)과 k-mean clustering이 있다[7]. 새로운 고객의 이웃 탐색을 위해 전체 사용자와 유사성 비교를 해야 하는 kNN방법과는 달리, k-mean clustering은 최초 클러스터를 생성한 이후에는 이웃 탐색을 위한 연산 비용을 줄일 수 있다[8].

협업필터링은 상품 평가 기록의 부족으로 인한 희박성(sparsity), 평가기록이 없는 새로운 고객과 상품에 대한 cold-start의 문제점을 갖고 있다[3][9]. 이밖에 본 연구의 사전실험 결과에 의해 협업필터링은 추천 성능 편차(inconsistent performance) 현상을 유발하는 것으로 확인되었으며 이 문제를 새로운 협업필터링의 특성으로 제기하고자 한다. 즉 협업필터링은 추천 대상자의 변동에 따라 상품 예측오차에 편차가 발생하며 비일관된 추천 성능을 보인다. 또한 단일 알고리즘일수록 이 현상이 심화되는 경향이 있다. 사전 실험으로 MovieLens 데이터셋에서 활동성이 높은 상위 100명의 사용자를 추출하여 순수 협업필터링(CF)과 사용자 클러스터링 CF(UC)와 아이템 클러스터링 CF(IC)을 적용하여 평균예측오차인 MAE(Mean Absolute Error)를 측정하였다. 실험결과에 다수 사용자에 대한 평균적인 MAE는 IC가 가장 우수하였으나 사용자별 최적의 알고리즘은 각각 달랐다. 그림 1의 그래프는 각 알고리즘별로 최소 MAE를 기록한 사용자의 비율을 표시한 것이다. IC가 46%의 사용자에 대하여 최소 MAE를 제공하였으나 54%의 사용자는 CF와 UC알고리즘을 적용하였을 때 예측오차를 더 줄일 수 있었다. 이러한 비일관된 예측오차의 편차는 기존 MAE 평가기준으로는 고려할 수 없으며 또 다른 지표인 정확도(precision)와 재현율(recall)로도 불가능하다.

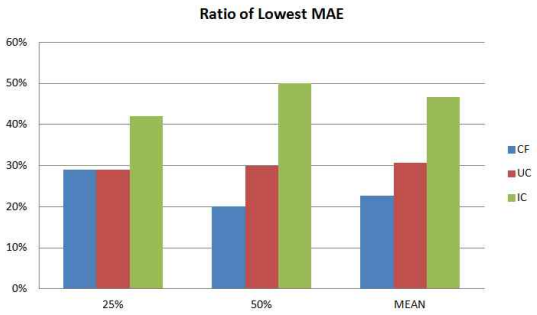

그림 1. 알고리즘별 최저 MAE 비율
Fig. 1. Low MAE ratio by Algorithm

협업필터링의 평가 방법에는 추천 정확도 측정 방법과 유용성 평가 방법이 있다. MAE, Precision, Recall등의 정확도 측정방법이 개발되었으나 이들을 포괄적으로 평가할 수 있는 연구는 상대적으로 부족하였다. 본 연구에서는 CF 알고리즘의 보다 정확한 성능 평가를 위하여 MAE편차를 비롯하여 MAE, 정확도와 재현율을 포괄하는 확장 성능평가모델로 MPR(MAE, Precision, Recall) 모델을 제안한다. 또한 단일 CF와 클러스터링 기반 복합CF에 대하여 제안된 확장 성능평가 모델을 적용하여 성능을 분석하는 실험을 수행한다.

2장에서는 협업필터링과 기존 평가방법을 소개하고, 3장과 4장에서는 클러스터링 기반 복합 협업필터링과 복합 평가모델인 MPR을 제안한다. 5장과 6장에서 실험 결과를 제시하고, 본 논문의 결론을 맺는다.

II. 관련 연구

개인화된 추천은 상품 검색에 소요되는 노력과 시간을 줄여 주어, 음악, 영화, 책, 앱과 같이 아이템의 수가 지속적으로 증가하는 전자상거래 시스템에서 유용하다[5]. 추천 시스템은 추천 알고리즘에 의해 그 성능이 좌우되며 크게 내용기반 추천, 협업필터링, 복합 알고리즘으로 구분된다[10-11].

1. 내용기반 추천

내용기반 추천 알고리즘은 사용자 프로필과 아이템의 속성정보 간의 연관성(나이, 성별, 위치, 관심분야, 구매내역 등)을 바탕으로 이루어진다[3-4][12]. 이 방법은 포트폴리오 효과(portfolio effect)를 유발하기 쉬운데, 즉 사용자가 이미 알고 있거나, 알고 있는 것과 유사한 아이템만을 주로 추천해 주는 것을 말한다 [11]. 또한 이 방법은 프로필 내용이 희박할 경우 추천의 질이 떨어질 수 있으며, 사용자와 아이템의 수가 증가할수록 알고리즘의 확장성이 떨어지게 된다[13].

2. 협업필터링

협업필터링(collaborative filtering: CF)[7][11][14-15]는 상업적으로 성공한 추천 알고리즘의 하나로 아이템 평가기록으로 부터 선호도가 유사한 사용자 그룹(이웃)을 탐색하고, 이 이웃들이 높게 평가한 아이템들 중 목표 사용자가 미경험한 아이템의 선호도를 예측하는 방법이다. 협업필터링은 표 1의 사용자-아이템 평가 행렬 상에서 수행되며, 여기에서 Rui는 사용자 u가 아이템 i에 대한 평가치(선호도)를 의미하고, n과 m은 각각 사용자와 아이템의 수이다.

표 1. 사용자-아이템 평가 행렬
Table 1. User-Item Rating Matrix

	$Item_1$	$Item_2$	…	$Item_m$
$User_1$	R_{11}	R_{12}	…	R_{1m}
$User_2$	R_{21}	R_{22}	…	R_{2m}
…	…	…	…	…
$User_n$	R_{n1}	R_{n2}	…	R_{nm}

내용기반 추천과는 대조적으로 협업필터링은 개별 아이템 간의 유사도에 의존하는 것이 아니라 사용자 평가 유사도에 기반하고 있기 때문에 여러 장르에 걸친(cross-genre), 예상하지 못한(serendipitous) 아이템들의 추천이 가능하다 [14]. 협업필터링의 처리과정은 다음과 같다.

선호도 유사도 계산(preference similarity)

협업필터링은 사용자-아이템 평가행렬을 사용하여 아이템에 대한 사용자의 선호도의 유사도를 계산하여 유사 선호도를 갖는 이웃(neighbor)을 탐색한다. 유사도 계산 방법에는 Pearson correlation coefficient, cosine similarity, mean square difference와 spearman correlation이 있으며, 이중 식(1)의 Pearson correlation coefficient가 가장 널리 사용된다 [1]. $r_{u,i}$ 는 사용자 u가 아이템 i를 평가한 값(rating)이고, $\bar{r_u}$ 는 사용자 u가 평가한 평가점수의 평균이다. Iu,k 는 사용자 u와 k가 공통으로 평가한 아이템의 집합을 의미하고, 식(1)을 통하여 유사도 상위 n명의 이웃이 결정된다.

$$sim(u,k) = \frac{\sum_{i \in I_{u,k}} (r_{u,i} - \bar{r_u})(r_{k,i} - \bar{r_k})}{\sqrt{\sum_{i \in I_{u,k}} (r_{u,i} - \bar{r_u})^2} \sqrt{\sum_{i \in I_{u,k}} (r_{k,i} - \bar{r_k})^2}} \quad (1)$$

평가점수 예측(preference prediction)

아이템 i에 대한 목표 사용자 u의 예측 평가치는 식(2)과 같이 계산된다. rv,i는 이웃 v가 아이템 i에 평가한 평가값을 의미하고 sim(u,v)는 사용자 u와 v간의 유사도이다. c는 사용자 u와 공통으로 아이템을 평가한 이웃의 수이고, 목표사용자와 이웃간의 유사도에 이웃의 과거 평가치를 결합하여 예상 선호도를 계산한다.

$$P_{u,i} = \bar{r_u} + \frac{\sum_{k=1}^{c} (r_{k,i} - \bar{r_k}) \times sim(u,k)}{\sum_{k=1}^{c} sim(u,k)} \quad (2)$$

3. 이웃탐색방법

k-Nearest Neighbor(kNN)[16] 탐색은 목표 사용자와 전체 사용자간의 선호 유사도를 계산한 후 유사도 상위 k명을 이웃으로 선정하는 방법이다. k가 너무 클 경우 추천 정확도나 성능을 떨어뜨릴 수 있으며, k가 너무 작으면 유사도가 높지 않은 사용자들의 아이템에 대해서는 예측할 수 없는 문제점이 발생하므로 이웃의 수를 적절하게 결정해야 한다.

클러스터링[7][8]은 유사 선호도를 갖는 사용자들의 클러스터를 생성하여 이웃으로 설정하는 방법이다. 클러스터링에는 평가한 상품들의 선호도가 유사한 고객들의 클러스터를 생성하는 사용자 클러스터링(user clustering)과 평가한 고객들이 유사하게 평가한 상품들의 클러스터를 생성하는 아이템 클러스터링(item clustering)이 있다[17][18]. 대표적인 k-mean

clustering[8]은 사용자의 선호도를 다차원 공간상의 점으로 표시하고 거리를 계산함으로써 전체 사용자의 집합을 여러 클러스터로 나누는 거리 기반 군집화 방법이다.

4. 협업필터링의 한계

협업필터링은 많은 장점을 제공하는 반면, cold-start, 확장성, 행렬의 희박성 등의 제약점이 존재한다[1][3][15].

Sparsity

일반적으로 고객들은 상품 평가를 잘 하지 않는 경향이 있으며, 평가행렬이 희박할 경우 추천의 질을 떨어뜨릴 수 있다. 특히 희박행렬은 몇몇 사용자들에 의해서만 높게 평가된 상품에 대해서는 유사한 사용자를 찾기 힘들기 때문에 추천이 어려울 수 있다[12]. 희박성 문제를 개선하기 위해, Pazzani[2]는 음식점 추천 시스템에서 성별, 나이, 지역코드, 학교, 직장 등의 인적 정보를 사용하였으며, Huang[15]은 사용자간 트랜지티브 연관성(transitive relevance)을 사용자 탐색 방법에 이용하였다.

Cold-start

평가를 전혀 하지 않은 새로운 사용자나 평가된 적이 없는 아이템은 협업필터링을 통해서 추천되기 어렵다. 대안으로 충분한 평가정보가 쌓일 때까지 사용자 프로필의 속성정보를 이용한 내용기반 추천을 병행할 수 있다[9].

Scalability

협업필터링은 사용자와 아이템의 수가 증가할수록 유사 이웃 탐색에 필요한 연산 비용도 비례하여 증가한다. 이는 실시간 추천을 어렵게 만드는 요인이 되며 확장성을 개선하기 위한 방법 중의 하나가 클러스터링 기법이다.

5. 협업필터링의 성능평가 방법

협업필터링의 성능평가는 추천 정확도(accuracy)를 측정하는 방법과 시스템의 유용성(usability)을 평가하는 방법으로 나눌 수 있다[19]. 정확도 측정방법은 표 2와 같이 예측정확도, 분류정확도, 순위정확도로 나뉜다.

표 2. 추천 정확도 측정 방법의 분류
Table 2. Classification of accuracy metrics

정확도의 분류	측정방법	측정도구
예측정확도 (Predictive accuracy)	예측 선호도와 실제 선호도의 차이를 측정	- MAE
분류 정확도 (Classification accuracy)	적절한 아이템인지 아닌지 올바르게 판단한 빈도	- Precision - Recall - F-measure
순위정확도 (Rank accuracy)	사용자와 추천 시스템이 아이템을 선호도에 따라 나열한 순위의 차이	- Item ranking

추천 시스템의 유용성 평가는 새롭게 부각되는 협업필터링의 평가방법으로 정확도 측정방법과 달리 추천 알고리즘이 포괄할 수 있는 아이템의 범위(coverage) 측정방법, 아이템 평가 자료의 수 대비 추천 질을 나타내는 학습율(learning rate) 평가방법, 인기 있는 제품 이외에 새롭고 참신한 상품을 추천할 수 있는 능력(novelty and serendipity) 평가방법, 추천 자신감(confidence) 평가 방법, 그리고 사용자 평가방법(user evaluation) 등이 있다[19].

III. 클러스터링 기반 복합 협업필터링

단일 협업필터링의 단점을 보완하기 위하여 인구통계학적(demographic) 정보를 활용하는 복합 CF에 대한 연구[4][9][12]가 이루어져 왔으나, 본 연구에서는 제안하는 확장 성능평가 모델을 적용하기 위하여 클러스터링 기반 복합 협업필터링을 설계한다.

1. 목표 사용자 그룹 추출

MovieLens 데이터셋(http://www.grouplens.org/)을 8:2의 비율로 train과 test 데이터셋으로 분리한 뒤, test 데이터셋에서 평가 횟수가 20이상으로 활동성이 높은 사용자들을 추출하여 목표 사용자로 설정한다.

2. 기본 CF 알고리즘 적용

성능평가를 위한 기본 알고리즘으로 순수 협업필터링(CF), 사용자 클러스터링 CF(User clustering CF: UC), 아이템 클러스터링 CF(Item clustering CF: IC)을 포함하며, 이를 목표 사용자에 각각 적용하여 1차 추천 리스트를 생성한다. 순수 협업필터링(CF)은 kNN 방법으로 선호도가 유사한 이웃 집단을 탐색하여 추천할 아이템의 선호도를 예측한다. UC는 그림 2의 K-mean 클러스터링을 이용하여 k개의 사용자 클러스터를 생성한다[17][20]. 이 방법은 이웃 탐색에 필요한 연산량을 줄여주는 반면, 적절한 군집의 수를 선택해야 하며 군집 중심의 초기값에 따라 수렴된 군집된 결과가 달라질 수 있다. 또한 서로 다른 클러스터에 속한 사용자들 간의 연관성을 고려할 수 없는 단점이 있다[7][8]. 클러스터링이 종료되면 선호도 예측은 클러스터내의 타 사용자들이 평가한 값의 평균으로 이루어

진다[17]. IC는 유사하게 평가된 아이템들의 클러스터를 생성한다는 점을 제외하고 UC와 동일하다 [18].

1. 전체 사용자 중 k명의 임의 사용자 선택하여 각 클러스트의 중심점으로 초기화한다.
2. 그다음 다른 모든 사용자들을 가장 거리가 가까운 클러스터에 할당한다. 거리계산은 Pearson Correlation coefficient나 Euclidean distance 방법을 이용하여 계산한다.
3. 모든 사용자의 클러스터 할당이 종료된 후, 각 클러스터에 속한 사용자들의 평가값을 평균하여 새로운 클러스터의 중심점을 계산한다.
4. 각 사용자와 새로운 클러스터 중심과의 거리를 계산하여 최단 거리의 클러스터에 재할당한다.
5. 소속 클러스터가 변경되는 사용자의 수가 임계치 범위 이내이면 클러스터링을 종료하고 그렇지 않으면 2번 과정부터 반복한다.

그림 2. K-mean 클러스터링
Fig. 2 K-mean Clustering

3. 추천 리스트 병합 알고리즘

복합 추천 리스트를 얻기 위하여 UC+IC와 CF+IC의 조합으로 1차 추천리스트들을 그림 3의 알고리즘을 통해 병합한다. 1개의 알고리즘에 의해서만 추천된 '미충돌 아이템'은 그대로 통합 리스트에 예측 점수와 함께 저장된다. 반면, 2개 이상의 알고리즘에 의해 각각 다른 예측 선호도로 중복 추천된 '충돌 아이템(collision items)'은 예측된 선호도 값들의 평균으로 최종 선호도를 결정한다.

```
Combined_algorithm ()
{
    Initialize Final_list
    For i <= number of CF algorithms do:
        While there is an item(j) to check in input_list(i)
            If item(j) is found in Final_list:
                Calculate the average of preference scores of them
                Assign the average as a new preference of the item
            Else if item(j) is not in Final_list
                Append item(j) into Final_list with its preference
        Endwhile
    Endfor
    Sort Final_list in descending order of item preference value
}
```

그림 3. 리스트 병합 알고리즘
Fig. 3 List Merge Algorithm

4. 최종 추천 리스트 생성

병합된 추천 리스트를 선호도 점수에 따라 내림차순으로 정렬하고 상위 m개의 아이템을 최종적으로 추천한다.

IV. 포괄적 성능평가 모델

기존의 협업필터링 성능평가 방법은 MAE를 중심으로 평균 예측 선호도 오차인 MAE를 중심으로 추천 정확도 평가[3]에 초점이 맞추어져 왔다. 비일관적인 예측성능편차 문제나 다른 지표와의 포괄적 성능평가에 대한 연구는 부족하였다. 기존의 성능평가 방법인 MAE 측정방법[3]은 사용자별 추천 오류편차는 무시하고 단순히 n명에 대한 MAE의 평균이 가장 낮은 알고리즘을 가장 우수한 것으로 판단한다. 또한 MAE와 추천 정확도, 재현율은 독립적으로 비교되어 왔으며 이들에 대한 종합적인 평가는 불가능하였다. 본 장에서는 MAE를 비롯한 기존 성능 지표를 포괄적으로 평가할 수 있는 확장 성능평가모델을 제시한다.

1. 확장 성능평가 모델

MAE는 CF에 의해 예측된 아이템 선호도와 실제 사용자의 평가치의 차를 평균한 것으로 식(3)과 같이 계산된다. Pi는 아이템 i에 대한 예측된 선호도이며 Ri는 사용자가 실제로 평가한 값이다. MAE가 낮을수록 알고리즘의 추천 오류가 적다는 것을 의미한다.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - R_i| \quad (3)$$

그러나 MAE는 평균적인 예측 선호도 오차를 의미하는 것으로 사용자별 MAE 편차는 반영하지 못하기 때문에, 최소 MAE를 갖는 알고리즘이 모든 사용자에게 최적의 알고리즘이 아닐 수 있다. 보다 정확한 CF알고리즘의 평가를 위하여 기존 지표인 MAE, MAE 편차, Precision, Recall을 포괄적으로 반영할 수 있는 확장형 성능평가 방법으로 MPRd(MAE, Precision, Recall) 모델을 제안한다. CF알고리즘의 안정성을 반영하고자 복수 CF 알고리즘의 상대적 성능 측정지표로 랭킹점수(Ranking Score: RS) 개념을 도입하였다. 즉 RS는 k명의 사용자에게 N개의 실험 알고리즘을 적용하여 MAE가 낮은 순으로 순위점수를 부여한 것으로 랭킹점수가 높을수록 알고리즘의 안정성이 높다.

$$RS = N - MR_i + 1 \quad (4)$$

여기서 N은 알고리즘의 수이고, MRi은 알고리즘 i의 MAE 순위이다. 즉, 현재 목표 사용자에 대한 MAE가 최소인 경우

최고점 N점을 받고, 최대인 경우 최저점 1점을 받는다.

식 (5)와 (6)은 CF알고리즘의 분류 정확도를 평가하기 위한 지표들이다. Precision은 추천된 아이템 중 몇 개가 고객이 실제로 좋아했던 것인지를 나타내는 비율이며, Recall은 고객이 좋아하는 아이템 중 얼마나 많은 아이템이 추천되었는지는 나타내는 비율이다[10].

$$Precision = \frac{|\{preferred\ items\}| \cap |\{predicted\ items\}|}{|\{predicted\ items\}|} \quad (5)$$

$$Recall = \frac{|\{preferred\ items\}| \cap |\{predicted\ items\}|}{|\{preferred\ items\}|} \quad (6)$$

식 (7)의 F-measure는 정확도와 재현율을 복합적으로 평가하는 기준이다[20].

$$F_{measure} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (7)$$

Precision과 Recall의 편차를 반영하고자 F-measure에 대해서도 RS를 측정한다. CF알고리즘의 추천 성능은 MAE값에는 반비례하고, RS, Precision, Recall, F-measure에 각각 비례한다고 볼 수 있다. 이를 정리하면, k명의 사용자에 대하여 N개의 알고리즘을 적용하였을 경우 MPRd은 식(8)로 정리된다. α는 F-measure와 MAE값의 범위를 보정하기 위한 계수이다.

$$MPR_d = \frac{\alpha \cdot F\_measure \cdot RS_{F\_measure}}{\frac{MAE}{RS_{MAE}}}$$

$$= \frac{\alpha \cdot F\_measure \cdot RS_{F\_measure} \cdot RS_{MAE}}{MAE} \quad (8)$$

where

$$\alpha = \frac{MEAN(\frac{MAE}{RS_{MAE}})}{MEAN(F\_measure \cdot RS_{F\_measure})}$$

2. 실험환경

성능평가 실험환경은 표3과 같으며, Movielens 데이터셋에서 추출된 목표사용자에 실험 알고리즘을 적용하여 알고리즘별로 각각 MAE, RS, Precision, Recall을 측정하고 F-measure, MPRd를 계산하였다.

표 3. 실험 데이터셋 및 측정값
Table 3. Experimental dataset and measurement

Category	Description
Train dataset	800,168 ratings (80%)
Test dataset	200,041 ratings (20%)
단일 CF	CF, UC, IC
복합 CF	UC+IC, CF+IC
Measured data	MAE, RS, Precision, Recall, F-measure, MPRd

그림 4는 MAE 측정 결과로 3개 비율의 평균 MAE에서 CF+IC(0.6776)가 가장 낮은 오차를 보여주었으며, 그 다음으로 UC+IC(0.6783)와 IC(0.6826)이었다. 복합 CF의 평균 MAE(0.6779)가 단일 CF(0.7009)보다 더 낮게 나타났다.

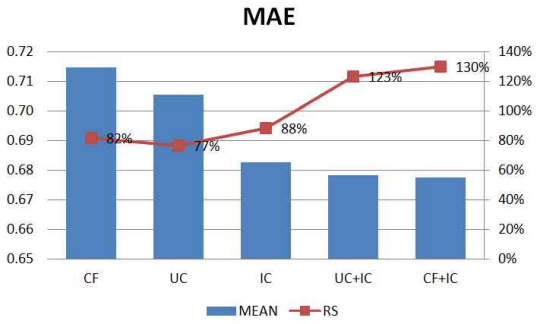

그림 4. MAE의 평균과 랭킹점수
Fig. 4. MEAN and RS of MAE

랭킹점수(RS)는 사용자에 대한 알고리즘의 안정성을 의미하며 MAE가 낮을수록 높은 점수가 부여된다. CF+IC가 가장 높은 점수를 기록했으며 UC+IC가 각각 그 다음 순위로 랭크되었다. 복합 CF 알고리즘의 평균 RS는 127%로 단일 CF 알고리즘보다 27%를 향상시킨 것으로 확인되었다. Fig. 5는 F-measure 측정값으로, UC(0.1496), UC+IC(0.1488)순으로 높았으며 F-measure랭킹점수는 UC+IC가 118%로 가장 높았다.

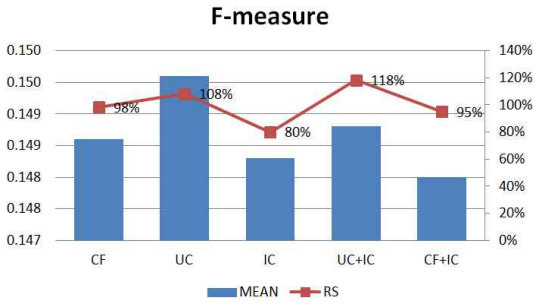

그림 5. F-measure의 평균과 랭킹점수
Fig. 5. MEAN and RS of F-measure

Fig. 6은 측정된 MAE, RS, F-measure를 반영하여 계산한

MPRd 값이다. 이 수치가 높을수록 CF 알고리즘의 포괄적 추천 성능이 우수하다는 것을 의미한다. MPRd지표에서 UC+IC 조합이 1.5409로 가장 우수한 추천 능력을 보여 주었다. 그 다음으로 CF+IC(1.2982)와 UC(0.8478)가 우수하였으며, CF와 IC가 가장 낮은 수치를 기록했다. 복합 알고리즘의 MPRd이 1.4196으로 단일 CF의 평균(0.7967)보다 높은 것으로 확인되었다.

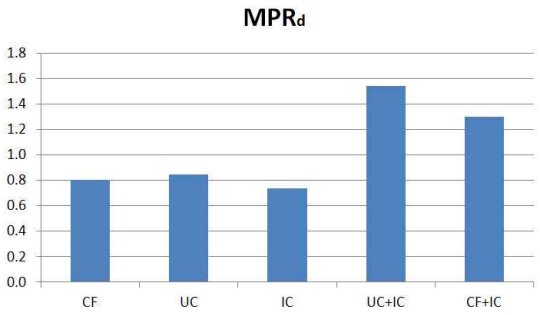

그림 6. MPRd
Fig. 6. MPRd

3. 토론

표 4는 평가항목별로 단일 CF 알고리즘 대비 복합 CF 알고리즘의 향상도를 계산한 것이다. MAE 항목에서 복합 CF은 단일 CF에 비하여 각각 3.3% 개선한 것으로 확인되었고, 특히 RS(MAE) 항목은 복합 CF 알고리즘에 의해 52.3%가 향상하였다. MPRd 항목에서는 78%가 개선된 것으로 계산되었다. q 본 연구의 가정과 같이 복합 CF는 포괄적 성능평가를 가능하게 하며 단일 알고리즘의 한계를 보완하여 전반적으로 우수한 성능을 제공하는 것을 알 수 있었다.

표 4. 단일 CF와 복합CF의 성능평가
Table 4. Performance comparison between single CF and combined CF

Category	단일 CF (S) (CF, UC, IC)	복합 CF (C) (UC+IC, CF+IC)	Improvement $\left(\frac{\lvert C-S\rvert}{S}\times 100\right)$
MAE	0.7009	0.6779	+3.3%
RS(MAE)	82%	127%	+54.1%
MPR_d	0.7968	1.4196	+78.1%

V. 결론 및 향후 연구

데이터의 희박성, Cold-start, 확장성은 CF 알고리즘의 근본적인 문제로 인식되어 지속적인 연구가 진행되고 있는 분야이다. 이밖에 협업필터링은 목표 사용자 변화에 대해 추천 성능의 편차가 발생하는 비일관성의 특성이 있음이 확인되었고, 본 연구에서는 CF알고리즘의 MAE와 추천편차를 포괄하는 MPRd 성능평가모델을 제시하였다. 성능평가결과, 복합 CF 알고리즘이 단일 CF 알고리즘에 비하여 전반적인 추천 성능과 안정성이 높은 것으로 확인되었다. 제안된 MPRd모델은 MAE 지표와 달리 수치의 직관성이 떨어지는 단점이 있으며, 개선된 평가모델을 향후 후속 연구주제로 제안하고자 한다.

참고문헌

[1] J. Konstan, D. B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying collaborative filtering to Usenet news," Communications of ACM, Vol. 40, No. 3, pp. 77-87, 1997.

[2] M. Pazzani, "A Framework for Collaborative, Content-Based, and Demographic Filtering," Artificial Intelligence Review, pp. 393-408, Dec 1999.

[3] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithm," Proc. of the 10th international conference on World Wide Web, pp. 285-295, 2001.

[4] A. Ferman, J. Errico, P. Beek, and M. Sezan, "Content-based Filtering and Personalization Using Structural Metadata," Proc. of the 2nd ACM/IEEE-CS Joint Conference on Digital libraries, NY, USA, pp. 393-393, 2002.

[5] H. Kwak,C. Lee,H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," Proc. the 19th International World Wide Web Conference, April 26-30, Raleigh NC USA, pp. 591-600, 2010.

[6] H. Liu and P. Maes, "InterestMap: Harvesting Social Network Profiles for Recommendations," Proc. of the Beyond Personalization Workshop,

San Diego, California, pp. 54-49, 2005.

[7] M. O. Connor and J. Herlocker, "Clustering Items for Collaborative Filtering," Proc. of the ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, 1999.

[8] C. Ding and X. He, "K-Means Clustering via Principal Component Analysis," Proc. of the 21th Int. Conf. on Machine Learning, pp. 225-232, 2004.

[9] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommendations," Proc. of the Eighteenth National Conference on Artificial Intelligence, Edmonton, Canada, pp. 187-192, July 2002.

[10] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering," Proc. of 22nd Annual International ACM SIGIR Conference, Research and Development in Information Retrieval, 1999.

[11] G. Groh and C. Ehmig, "Recommendations in Taste Related Domains: Collaborative filtering vs. Social filtering," Proc. of GROUP'07, pp. 127-136, 2007.

[12] M. Balabanovic and Y. Shoham, "Fab: Content-Based, Collaborative Recommendation," Communications of the ACM, Vol. 40, No. 3, pp. 66-72, 1997.

[13] Y. Yang and J. Li, "Interest-based Recommendation in Digital Library," Journal of Computer Science, Vol. 1, No. 1, pp. 40-46, 2005.

[14] J. L. Herlocker, J. A. Konstan, and J. T. Riedl, "An Empirical Analysis of Design Choices in Neighborhood-based Collaborative Filtering Systems," Information Retrieval, Vol. 5, pp. 287 - 310, 2002.

[15] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," ACM Trans. Information Systems, Vol. 22, No. 1, pp. 116- 142, 2004.

[16] S. Brin, "Near Neighbor Search in Large Metric Spaces," Proc. of the 21th International Conference on Very Large DataBases, pp. 574-584, 1995.

[17] F. Zhang, H. Liu, and J. Chao, "A Two-stage Recommendation Algorithm Based on K-means Clustering In Mobile E-commerce," Journal of Computational Information Systems, Vol. 6, No. 10, pp. 3327-3334, 2010.

[18] S. Gong, "A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering," Journal of Software, Vol. 5, No. 7, July 2010.

[19] J. Herlocker, J. Konstan, L. Terveen, and J. Riedl, "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, Vol. 22, No. 1, pp. 5 - 53, January 200.

[20] T. Kim, S. Park, and S. Yang, "Improving Prediction Quality in Collaborative Filtering based on Clustering," Proc. of 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 704-710, 2008.

저 자 소 개

유 석 종

1994: 연세대학교 컴퓨터과학과 이학사
1996: 연세대학교 컴퓨터과학과 이학석사
2001: 연세대학교 컴퓨터과학과 공학박사
현재 : 숙명여대 컴퓨터과학부 부교수
관심분야 : 소셜컴퓨팅, 추천시스템
Email : sjyu@sm.ac.kr