

# 추천시스템의 희소성이 예측 정확도에 미치는 영향에 관한 연구

## The Effect of Data Sparsity on Prediction Accuracy in Recommender System

김 선 옥\*      이 석 준\*\*  
Kim Sun-ok    Lee Seok-jun

### 요 약

협력적 필터링을 이용한 추천시스템은 희소성의 문제로 인해 예측의 정확도에 대한 신뢰성에 문제가 있다. 이는 선호도 평가치의 희소성이 크면 이웃선정과정에서 문제가 있을 뿐만 아니라 예측의 정확도를 떨어뜨린다. 본 논문에서는 사용자의 응답 희소성에 따른 MAE의 변화를 조사하였으며 희소성에 따라 집단을 분류하고 분류된 집단에 따른 MAE는 유의적인 차이가 있는지를 분석하였다. 그리고 희소성 문제로 인한 집단 간의 예측 정확도를 높이기 위한 방법으로 희소성이 있는 아이템을 선별하여 이들 중에서 선호도 응답이 많은 사용자 고객의 선호도 평균값을 선호도 평가치로 대체시켜 희소성을 완화하여 추천시스템의 예측 정확도가 높아졌음을 연구하였다.

### Abstract

Recommender System based on the Collaborative Filtering has a problem of trust of the prediction accuracy because of its problem of sparsity. If the sparsity of a preference value is large, it causes a problem on a process of a choice of neighbors and also lowers the prediction accuracy. In this article, a change of MAE based on the sparsity is studied, groups are classified by sparsity and then, the significant difference among MAEs of classified groups is analyzed. To improve the accuracy of prediction among groups by the problem of sparsity, We studied the improvement of an accurate prediction for recommending system through reducing sparsity by sorting sparsity items, and replacing the average preference among them that has a lot of respondents with the preference evaluation value.

□ Keyword : Recommender System, Sparsity, Mean Absolute Error, Collaborative Filtering

## 1. 서론

인터넷의 보급과 네트워크의 발전에 따라 정보의 양이 폭발적으로 넘쳐나고 있다. 사용자에게 사용자가 원하는 정보를 알맞게 선별해 주는 기능은 사용자에게 큰 도움이 된다[1]. 또한 개인화 서비스의 방법으로 사용자가 원하는 정보를 제공할 수 있는 기능은 매우 중요하다. 개인화 서비스

의 자동화 기능을 가지고 있는 추천시스템은 사용자의 각각에 대하여 개인에게 적합한 정보나 서비스를 개별적으로 제공해 주는 특징이 있다 [2,3]. 또한 전자상거래를 통한 구매촉진에 기여하기도 하여 Amazon.com과 Moviercritic.com 등과 같은 일부 사이트들은 추천시스템을 이용하여 서비스를 제공하고 있다. 여기서 말하는 추천시스템이란 방대한 양의 사용자와 아이템에 관한 정보를 명시적 또는 묵시적으로 얻어서 조사하고 분석하여 이들 관계를 정확히 파악하여 사용자가 좋아할 만한 아이템을 추천해 주는 시스템이다. 추천시스템은 크게 내용기반 필터링과 협력적 필터링

\* 정 회 원 : 한라대학교 정보통신공학부 전임강사  
sokim@halla.ac.kr

\*\* 정 회 원 : 상지대학교 경영학과 겸임교수  
digitaldesign@sangji.ac.kr(교신저자)

[2007/08/20 투고 - 2007/08/27 심사 - 2007/09/19 심사완료]

의 두 가지 방법으로 분류될 수 있다. 본 논문은 협력적 필터링 기법에 대하여 연구하였다. 협력적 필터링은 사용자로 하여금 잘 정의된 영역에서 자신이 좋아하는 영역을 선택하거나 우선순위를 부여하여 그 정보를 바탕으로 비슷한 성향을 가진 이웃의 정보를 이용하여 추천하는 시스템이다 [4,5]. 그러나 협력적 필터링을 이용한 추천시스템은 일정 수 이상의 평가치가 있어야 하며 사용자의 평가가 적은 경우 희소성으로 인해 예측의 정확도가 낮아지는 단점이 있다. 희소성의 문제는 사용자의 평가 자료가 전체 평가 자료에 비해 현저히 부족하여 이웃과의 관계를 나타내는 유사도를 계산 할 경우 두 사용자의 관계를 정확하게 파악하지 못해 예측의 정확성을 떨어뜨리는 단점을 가지고 있어 추천시스템의 신뢰도를 떨어뜨리는 문제가 발생된다. 이러한 희소성은 시스템의 초기 구축 단계에서 발생할 수 있는 문제점이다. 본 논문에서는 희소성이 추천시스템에 미치는 영향에 대해 분석하였으며 희소성이 있는 아이템을 선별하여 이 아이템에 대한 희소성 완화를 통하여 추천 정확도를 향상시켜 추천시스템의 신뢰도를 높이고자 한다.

## 2. 관련연구

필터링을 이용한 추천시스템에 대한 연구는 Rich[6]가 제안한 초기부터 컴퓨터가 각 가정에 보급되고 인터넷이 발달됨에 따라 강력한 시스템으로 인식되어 왔다. 1992년 Goldberg[7]는 “협력적 필터링”이라는 용어를 도입하여 문맥기반의 시스템에서 이메일을 필터링하기 위해 처음으로 사용하였다. Shardanand[8]은 음악을 추천하기 위해서 협력적 필터링을 이용하였다. GroupLens[9]는 협력적 필터링에서 사용자의 선호도를 알아내기 위해 상관계수를 사용하게 되었고, 벡터 유사도 가중치를 이용하여 추천시스템의 성능 향상을 추진하였다. Pazzani[10]는 희소성이 있는 데이터가 있을 때 사용자의 선호도 예측을 위하여 속성

추출의 기술을 제안하였고, 사용자-아이템에 선호도 행렬의 차원을 줄이기 위해 SVD(Singular Value Decomposition)을 사용하였다. Breese[11]는 사용자로부터 평가 데이터가 없는 희소성의 문제를 해결하기 위한 방법으로 정보가 없는 아이템은 기본값을 사용하고, 각 아이템에 대해 사용자의 평가 정도에 따라 서로 다른 가중치를 적용하였다. Herlocker[12]는 다양한 유사도와 유사도 가중치 방법에 대한 연구로 희소성의 문제를 해결하였다. Hyungil[13]는 희소성의 문제를 해결하기 위하여 추가 속성 정보의 확률분포를 이용한 데이터 블러링 기법을 사용하였다. EunJu[14]은 희소성문제의 해결 방안으로 신경망을 이용하여 사용자와 아이템들 간의 선호도 상관관계를 학습시켜 모델을 생성하고 생성된 모델을 이용하여 추천 아이템의 선호도를 추천하였다. 협력적 필터링에 의한 추천시스템은 유사한 사용자 평가를 기반으로 아이템을 추천하기에 내용기반 필터링에 비해 정확하며, 아이템 자체에 대한 내용정보가 없어도 그 아이템에 대한 추천이 가능하다는 장점이 있다. 하지만 협력적 필터링에 대한 다양한 연구가 진행되고 있지만 협력적 필터링은 아직까지 데이터의 희소성과 초기 추천에 문제를 지니고 있다. 본 논문에서는 희소성 문제를 해결하기 위해 아이템에 대한 희소성을 조사하여 협력적 필터링에서 희소성 아이템이 MAE에 미치는 영향을 분석하였다. 그리고 희소성 문제로 MAE를 완화하기 위한 방법으로 희소성이 있는 아이템을 선별하고 선별된 아이템 중 선호도 응답이 많은 사용자 고객의 선호도 평균값을 선호도 평가치로 대치하여 MAE가 개선됨을 연구하였다.

### 2.1 협력적 필터링

협력적 필터링은 아이템의 특성이나 사용자의 프로파일과 같은 속성은 의도적으로 무시하고 사용자들이 아이템에 대해 평가한 선호도 즉 사용자-아이템 간의 관계만을 이용하여 선호도를 예

측하는 시스템이다. 따라서 협력적 필터링은 아이템의 내용 정보를 필요로 하지 않기 때문에 내용을 분석하기 어려운 음악이나 영화 같은 아이템 추천에 많이 사용되고 있다. 협력적 필터링에 의한 추천시스템은 아이템에 대한 각 사용자들의 평가 정보를 이용한다[2,3,4]. 가장 일반적인 접근 방법은 사용자들 사이의 평가 정보를 비교하여 유사 사용자를 추출하고, 아이템에 대한 유사 사용자의 선호도를 기반으로 특정 아이템에 대한 사용자의 선호도를 예측하는 것이다. 본 논문에서는 유사 선호도 사용자를 추출하기 위해 피어슨의 상관계수를 사용하고 이웃기반 협력적 필터링을 이용하여 선호도 예측 값을 산출하여 추천 대상 고객에게 제시할 추천 대상 아이템을 추출한다.

## 2.2 이웃기반 협력적 필터링 알고리즘

이웃기반 협력적 필터링을 사용한 추천시스템은 GroupLens에서 제시되었으며 추천 대상 고객의 선호도뿐만 아니라 이웃 고객들의 선호도도 함께 이용하여 선호도를 예측하게 된다 (Resnick et al., 1994). 다음 식은 GroupLens에서 제안한 이웃기반 협력적 필터링(Neighborhood Based Collaborative Filtering Algorithm)을 이용한 선호도 예측값에 대한 알고리즘이다.

$$\hat{U}_x = \bar{U} + \frac{\sum_{j \in \text{Raters}} (J_x - \bar{J}) r_{uj}}{\sum_{j \in \text{Raters}} |r_{uj}|}, \quad \bar{J} = \frac{\sum_{i=1}^n J_i}{n}, \quad i \neq x \quad (1)$$

여기서  $\hat{U}_x$ 는 아이템  $x$ 에 대한 추천 대상 고객  $u$ 의 선호도 예측치이며  $\bar{U}$ 는 추천 대상 고객  $u$ 가 평가한 모든 상품에 대한 평균이다.  $J_x$ 는 아이템  $x$ 에 대한 이웃 고객  $j$ 의 선호도 평가치이고,  $\bar{J}$ 는 이웃 고객  $j$ 가 평가한 모든 상품에 대한 선호도의 평균이다.  $\bar{J}$ 의 값을 정할 때 평가치 중에서 아이템  $x$ 에 대한 평가치는 제외한다.  $r_{uj}$ 는 추천 대상 고객  $u$ 와 추천 대상 고객의 이웃

고객인  $j$ 의 선호 유사 정도를 나타내는 유사도 가중치이다. 유사도 가중치는 여러 형태로 정의될 수 있으며 피어슨 상관계수가 많이 사용되고 있으며 본 연구에서도 피어슨 상관계수를 유사도 가중치로 이용한다. 피어슨 상관계수에 대한 정의는 다음과 같다.

$$r_{uj} = \frac{\sum_{i=1}^m (R_{u,i} - \bar{R}_u)(R_{j,i} - \bar{R}_j)}{\sqrt{\sum_{i=1}^m (R_{u,i} - \bar{R}_u)^2 \cdot \sum_{i=1}^m (R_{j,i} - \bar{R}_j)^2}} \quad (2)$$

여기서,  $r_{uj}$ 는 선호도를 예측할 고객  $u$ 와 이웃 고객  $j$ 와의 유사도 가중치이며  $R_{u,i}$ 는 추천 대상 고객  $u$ 가 평가한  $i$ 번째 아이템에 대한 선호도 평가치이고  $\bar{R}_u$ 는 추천 대상 고객  $u$ 가 평가한 아이템들에 대한 평균이다. 단, 유사도 가중치 계산을 위해 포함되는 평가치는 추천 대상 고객  $u$ 와 이웃고객  $j$ 가 공통으로 평가한 평가치로만 계산된다.

## 3. 선호도 예측의 성능 평가

협력적 필터링에서 예측의 정확도를 평가하기 위한 알고리즘의 성능 정확도는 실제 선호도와 협력적 필터링을 통한 예측 선호도의 절대평균오차(Mean Absolute Error)를 이용한다. 따라서 MAE(Mean Absolute Error)가 작으면 예측 정확도는 높아진다(Shardanand and Maes, 1995). 다음 식은 예측의 정확도를 평가하기 위한 MAE의 계산식이다.

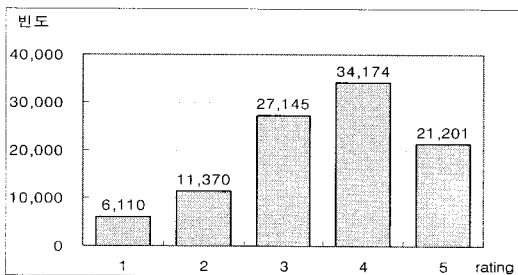
$$MAE = \frac{1}{N} \sum_{j=1}^N |R_{uj} - \hat{R}_{uj}| \quad (3)$$

여기서,  $R_{uj}$ 는 아이템  $j$ 에 대한 추천 대상 고객  $u$ 의 실제 선호도 평가치이고  $\hat{R}_{uj}$ 는 아이템  $j$ 에 대한 추천 대상 고객  $u$ 의 선호도 평가치에 대한 예측 값이다.

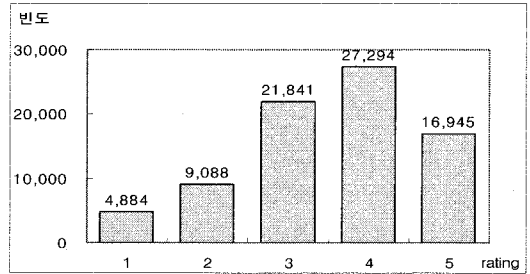
## 4. 실험

### 4.1 실험데이터

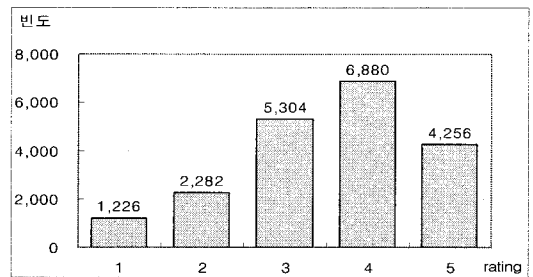
본 논문에서는 GroupLens에서 제공되는 MovieLens 100k 자료를 실험데이터로 사용하였다. MovieLens 100k 자료는 943명의 사용자가 1682편의 영화를 평가한 정보가 100,000개 포함되어 있으며 사용자는 적어도 20편의 영화에 대한 선호도를 평가하도록 하였다. 사용자는 1에서 5점의 범위로 평가하게 되어있다. 본 논문에서는 GroupLens에서 제공되는 MovieLens 100K dataset을 80%의 training dataset과 20%의 test dataset으로 랜덤하게 분할하였으며 training dataset과 test dataset은 각각 2개의 dataset으로 만들어 특정한 dataset에서 발생할 수 있는 특정한 결과의 오류를 배제하였다. 2개의 dataset인 dataset1(training dataset, test dataset), dataset2(training dataset, test dataset)는 유사한 형태의 빈도분포를 나타내었다. 그림1,그림2,그림3은 MovieLens의 100K dataset 전체에 대한 선호도 평가치 빈도분포와 dataset1의 training dataset 그리고 test dataset에 대한 선호도 평가치의 빈도 분포이다. dataset2도 유사한 형태의 빈도분포로 나타났다. 전체 자료와 비교할 때 training dataset과 test dataset의 빈도분포형태는 유사한 모양의 분포를 하고 있어 실험의 적용에 적합하다고 판단되었다.



(그림 1) 전체 dataset에 대한 빈도 분포



(그림 2) training dataset에 대한 빈도 분포



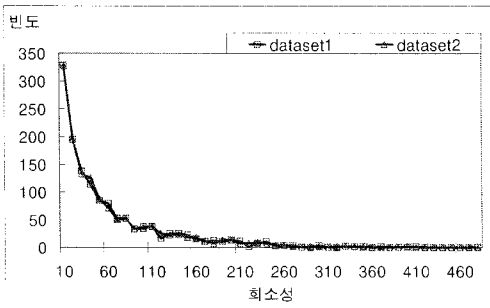
(그림 3) test dataset에 대한 빈도 분포

### 4.2 실험방법

본 연구는 GroupLens의 MovieLens dataset의 training dataset의 훈련을 통해 test dataset의 예측은 user based로 진행하였다. training dataset에서 희소성이 있을 경우 test dataset의 아이템에 대한 예측치의 정확도를 MAE로 이용하여 계산하고 희소성이 큰 경우와 희소성이 작은 경우로 나누어 MAE의 유의적인 차이가 있는 지를 분석하였다. 희소성은 응답의 평가치가 작은 경우에 발생하는 것이므로 희소성의 크기를 변화시키면서 MAE의 차이가 어떻게 나타나는지를 분석하였다. 예측 결과 1682개의 아이템 중에서 이웃 선정이 되지 않아 선호도 예측을 할 수 없는 아이템을 제외하고 분석하였다. 또한 희소성이 어떤 영향을 미치는지를 분석하기 위해 임의로 정한 희소성의 수보다 적은 집단을 집단1, 희소성의 수보다 응답수가 많은 집단을 집단2로 분류하여 희소성이 MAE에 미치는 변화를 분석하였다.

실험에 사용된 dataset1과 dataset2를 희소성의

수에 따른 빈도는 그림4와 같다.

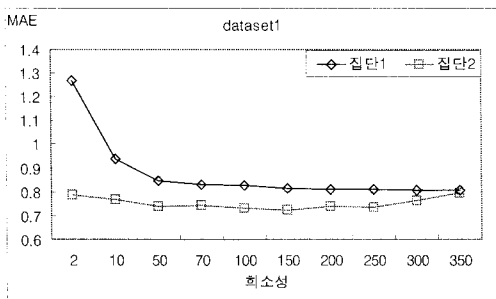


(그림 4) training dataset의 희소성의 수의 빈도

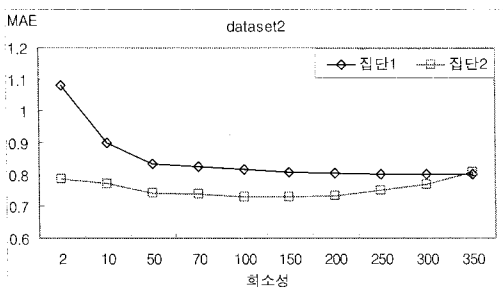
그림4에서 dataset1, dataset2에서 training dataset의 희소성의 빈도는 유사하게 나타났다.

### 4.3 실험결과

그림5, 그림6은 희소성의 수에 따른 MAE의 변화량이다.



(그림 5) dataset1의 training dataset의 희소성에 따른 MAE



(그림 6) dataset2의 희소성에 따른 MAE

dataset1, dataset2는 희소성에 따라 집단 간의 MAE 차이가 있음을 보여주고 있다. 희소성이 클수록 두 집단 간 MAE의 차이가 크며 희소성이 작을수록 두 집단 간 MAE의 차이가 작아지는 것으로 나타났다. 표1, 표2는 training dataset에서 희소성에 따라 분류된 집단의 test dataset에서 MAE 평균의 차이가 있는지를 알아보기 위하여 training dataset의 두 집단에 대한 test dataset의 MAE의 t-test 검정결과이다.

(표 1) dataset1의 training dataset에서 희소성에 따른 test dataset의 MAE의 t-test 검정결과

| 희소성 | MAE   |       | t값    | 유의확률    |
|-----|-------|-------|-------|---------|
|     | 1집단   | 2집단   |       |         |
| 2   | 1.270 | 0.787 | 4.310 | 0.000** |
| 10  | 0.940 | 0.766 | 4.900 | 0.000** |
| 50  | 0.847 | 0.741 | 6.180 | 0.000** |
| 70  | 0.832 | 0.743 | 5.660 | 0.000** |
| 100 | 0.825 | 0.733 | 6.160 | 0.000** |
| 150 | 0.816 | 0.724 | 6.050 | 0.000** |
| 200 | 0.811 | 0.738 | 4.070 | 0.000** |
| 250 | 0.809 | 0.737 | 3.210 | 0.002** |
| 300 | 0.808 | 0.764 | 1.500 | 0.151   |
| 350 | 0.808 | 0.796 | 0.310 | 0.767   |
| 400 | 0.808 | 0.767 | 0.210 | 0.834   |

\* :  $p < 0.05$ , \*\* :  $p < 0.01$

(표 2) dataset2의 training dataset에서 희소성에 따른 test dataset의 MAE의 t-test 검정결과

| 희소성 | MAE   |       | t값     | 유의확률    |
|-----|-------|-------|--------|---------|
|     | 1집단   | 2집단   |        |         |
| 2   | 1.080 | 0.788 | 2.740  | 0.008** |
| 10  | 0.899 | 0.771 | 3.610  | 0.000** |
| 50  | 0.836 | 0.742 | 5.590  | 0.000** |
| 70  | 0.826 | 0.739 | 5.580  | 0.000** |
| 100 | 0.818 | 0.731 | 5.960  | 0.000** |
| 150 | 0.809 | 0.732 | 5.090  | 0.000** |
| 200 | 0.805 | 0.734 | 4.080  | 0.000** |
| 250 | 0.803 | 0.752 | 2.040  | 0.048*  |
| 300 | 0.802 | 0.768 | 1.120  | 0.278   |
| 350 | 0.802 | 0.809 | -0.170 | 0.872   |
| 400 | 0.802 | 0.759 | 1.230  | 0.325   |

\* :  $p < 0.05$ , \*\* :  $p < 0.01$

표1, 표2의 t-test 검정결과 희소성이 크면(응답수가 작으면) 집단 간에는 유의적인 차이가 있는 것으로 나타났으며 희소성이 작으면(응답수가 크면) 유의적인 차이가 나타나지 않는 것으로 나타났다. 또한 희소성이 크면 MAE가 크게 나타나는 것으로 나타났다. 이것은 희소성이 MAE에 영향을 미친다고 할 수 있다. 즉 희소성은 추천시스템의 예측정확도를 저하시킨다고 할 수 있다. dataset2에서도 유사한 결과를 보이고 있다. 따라서 희소성을 완화하기 위한 방법으로 본 논문에서는 이러한 희소성 문제로 인한 예측 정확도의 영향을 개선하기 위해 희소성이 있는 아이템을 선별하고 선별된 아이템중 선호도 응답이 많은 사용자 고객의 선호도의 평균값을 선호도 평가치로 대치하여 MAE에 대한 변화를 비교하였다. 먼저, dataset1과 dataset2에서 희소성이 큰 training dataset의 아이템에 사용자 평가치의 평균을 대치시킨 후 user based에 의한 협력적 필터링 방법으로 test dataset에 대해 예측을 한 후 MAE에 대한 변화를 비교하였다. 여기서, dataset1을 두개의 dataset11, dataset12로 분류하고 희소성이 큰 아이템을 선정하여 dataset11에는 9개의 아이템을 추가하고 dataset12에는 17개의 아이템을 추가하였다. 마찬가지로 dataset2에서도 dataset21에는 희소성이 큰 아이템을 4개 추가하였고 dataset22에는 17개를 추가하였다. 표3은 training dataset의 희소성이 있어 선정된 아이템에 사용자 평가치의 평균을 각각 10개, 15개 대치시킨 후 MAE가 개선되었는지를 분석한 결과이다.

(표 3) dataset에서 희소성 아이템 수에 따라 분류된 MAE의 t-test 검정결과

| dataset   | 대응     | N    | 평균     | 평균차    | t값    | 유의확률    |
|-----------|--------|------|--------|--------|-------|---------|
| dataset11 | 무조건    | 1378 | 0.8079 | 0.0085 | 2.692 | 0.003** |
|           | 10개 추가 |      | 0.7994 |        |       |         |
| dataset12 | 무조건    | 1378 | 0.8079 | 0.0101 | 2.764 | 0.002** |
|           | 15개 추가 |      | 0.7978 |        |       |         |
| dataset21 | 무조건    | 1379 | 0.8021 | 0.0043 | 1.959 | 0.025*  |
|           | 10개 추가 |      | 0.7977 |        |       |         |
| dataset22 | 무조건    | 1379 | 0.8021 | 0.0106 | 3.205 | 0.000** |
|           | 15개 추가 |      | 0.7914 |        |       |         |

\* : p<0.05 , \*\* : p<0.01

모든 dataset에서 희소성 아이템의 수를 선정하여 검정한 결과는 예측의 정확도에 대한 MAE가 낮아짐을 알 수 있다. 특히, 희소성 아이템의 수가 17개인 dataset22는 희소성을 완화했을 경우에 MAE의 평균차이가 높게 나타났다.

## 5. 결론

협력적 필터링을 이용한 추천시스템은 데이터의 희소성이 있을 때 예측의 정확도가 떨어진다는 단점을 가지고 있다. 본 논문에서는 협력적 필터링에서 희소성 문제에 대하여 사용자의 이웃기반 협력적 필터링을 이용하여 아이템에 대한 정보를 기반으로 희소성에 따라 MAE 값을 비교한 결과 예측정확도에 차이가 있음을 분석하였다. 희소성에 따라 분류된 집단에서 각 집단의 희소성이 클수록 추천시스템의 예측 정확도가 떨어져 추천시스템의 신뢰도가 떨어질 수 있다는 것을 알 수 있었으며 희소성을 완화했을 경우에 MAE가 개선됨을 알 수 있었다. 차후 연구로 희소성이 있을 때 희소성을 완화하고 MAE를 개선하기 위하여 어떤 특징이 있는 자료를 대치하는 것이 MAE를 줄일 수 있는지에 대한 연구가 필요하며 user based에 의한 대치방법과 item based의 대치방법을 서로 비교연구 할 필요가 있다.

## 참고 문헌

- [1] 한정혜, 변루나, "실시간 추천을 위한 분할셋 기반 Up-to-Moment 선호모델 탐색", 인터넷 정보학 논문지, Vol.8, No.2, pp.105-115, 2007.
- [2] 이수봉, 이재완, "연관규칙을 이용한 전자상거래 PG솔루션", 한국 인터넷 정보학회 2004 춘계학술발표대회 논문집, 제5권, 제1호, pp.71-75, 2004.
- [3] 이승환, 빈병기, 최동운, "퍼지 이론을 이용한 맞춤형 쇼핑물을 위한 지능형 에이전트", 한국 인터넷 정보학회 춘계학술발표대회 논

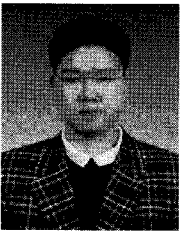
- 문집, 제2권 1호, pp.375-381, 2001.
- [4] 이희춘, 이석준, “사용자 기반 추천시스템에서 근접이웃 알고리즘과 수정알고리즘의 예측 정확도에 관한 연구”, Journal of the Korean Data Analysis Society, Vol.8, No.5, pp.1893-1904, 2006.
- [5] Lee, H. C., S. J. Lee, Y. J. Chung, “The Effect of Co-rating on the Recommender System of User Base”, Journal of the Korean Data & Information Society, Vol.17, No.3, pp.775-784, 2006.
- [6] Elaine Rich, “User modeling via stereotypes”, Cognitive Science, Vol.3, pp.335-366, 1979.
- [7] David Goldberg, David Nichols, Brian Oki, and Douglas Terry, “Using collaborative filtering to weave an information tapestry”, Communication of the ACM, Vol.35, No.12, pp.61-70, 1992.
- [8] U.Shardanand and P.Maes. “Social information filtering : Algorithms for automating word of mouth”, In ACM Conference on Computer Human Interaction(CHI), 1995.
- [9] J. Konstan, B. Miller, D.Maltz, J. Herlocker, L. Gordon, and J. Riedl, “GroupLens: Applying Collaborative Filtering to Usenet News”, Communications of the ACM, Vol.40, No.3, pp.77-87, 1997.
- [10] Pazzani, M.J., “A Framework for Collaborative, Content\_Based and Demographic Filtering”, Artificial Intelligent Review, pp.394-408, 1999.
- [11] Breese, Heckerman, and Kadie. “Empirical analysis of predictive algorithms for collaborative filtering”, Microsoft Research Technical Report, (MSR- TR -98 -12), October 1998.
- [12] Jonathan Herlocker, Joseph Konstan, Al Borchers, and John Riedl. “An algorithmic framework for performing collaborative filtering”, In Proceedings of the SIGIR. ACM, August 1999.
- [13] Hyungik Kim, Juntae Kim. “Data Blurring Method for Solving Sparseness Problem in collaborative filtering”, Journal of KISS, Vol.32, No.6, pp.542-553, 2005.
- [14] Eun Ju Kim, Joung Woo Ryu, Myung Won Kim. “Collaborative filtering for Recommendation based on Neural Network”, Journal of KISS, Vol.31, No.4, pp.457-466, 2004.
- [15] Resnick, P., N. Iacovou, M. Suchak, P. Bergstrom, J. Riedl, “GroupLens: an open architecture for collaborative filtering of netnews”, in Proceedings of the 1994 ACM conference on Computer supported cooperative work. ACM Press: Chapel Hill, North Carolina, United States. pp. 175-186, 1994.

● 저자 소개 ●



**김 선 옥(Kim Sun-ok)**

1991년 서강대학교 대학원 수학과 졸업(석사)  
1998년 서강대학교 대학원 수학과 졸업(박사)  
1998년 -2004년 상지대, 서강대, 한라대 시간강사  
2005년-현재 한라대학교 정보통신공학부 전임강사  
관심분야 : 정보보안, 전자상거래, 추천시스템, 암호학 etc.  
E-mail sokim@halla.ac.kr



**이 석 준(Lee Seok-jun)**

1996년 상지대학교 산업공학과 졸업(학사)  
1999년 상지대학교 산업환경대학원 산업공학과 졸업(석사)  
2007년 상지대학교 대학원 경영학과 졸업(박사)  
2006년~현재 상지대학 경영학과 겸임교수  
관심분야 : 전자상거래, 추천시스템, 데이터마이닝, 데이터베이스, etc.  
E-mail : digitaldesign@sangji.ac.kr