

## Pre-Evaluation for Detecting Abnormal Users in Recommender System

Seok-Jun Lee<sup>1)</sup> · Sun-Ok Kim<sup>2)</sup> · Hee-Choon Lee<sup>3)</sup>

### Abstract

This study is devoted to suggesting the norm of detection abnormal users who are inferior to the other users in the recommender system compared with estimation accuracy. To select the abnormal users, we propose the pre-filtering method by using the preference ratings to the item rated by users. In this study, the experimental result shows the possibility of detecting the abnormal users before the process of preference estimation through the prediction algorithm. And It will be possible to improve the performance of the recommender system by using this detecting norm.

**Keywords** : Collaborative Filtering, Recommender System

### 1.서론

전자상거래의 확대는 전통적인 판매와 마케팅 전략에 큰 변화를 가져오고 있다. 특히 전자상거래 사이트에서 거래되는 상품은 오프라인 매장에서 진열할 수 없을 만큼 다양하고도 수많은 상품들로 이루어져 있으며 거래 상품도 유형의 상품뿐만 아니라 정보서비스 및 콘텐츠 부문으로 크게 확대되고 있다. 그렇기 때문에 전자상거래에서 상품 판매 전략은 기존의 판매 전략과는 차별화된 전략을 채택하여야만 한다. 또한 오프라인 매장에서의 상품 판매는 고객과 직원의 1:1 관계에서 거래가 성립되지만 전자상거래의 경우 시간과 공간의 제약뿐만 아니라 고객과의 관계에 있어서의 제약도 허물고 있다. 그러나 전자상거래에서 거래되는 상품의 종류와 양이 증가함에 따라 이를 이용하는 고객에게는 새로운 정보과잉에 따른 문제점을 부여하고 있다. 정보과잉

---

1) 강원도 원주시 우산동 660번지 상지대학교 경영학과 겸임교수  
E-mail : crco909@yahoo.co.kr

2) 강원도 원주시 한라대1길 32 한라대학교 정보통신공학부 교수  
E-mail : sokim@halla.ac.kr

3) 교신저자: 강원도 원주시 우산동 660번지 상지대학교 컴퓨터 데이터 정보학과 교수  
E-mail : choolee@sangji.ac.kr

의 문제는 고객들의 정보검색비용을 증가시키게 되며 정보검색비용의 증가는 곧 고객의 만족도를 저하시키는 원인이 될 수 있다. 이러한 정보과잉에 따른 고객의 만족도 저하를 해결하기 위한 접근법으로 추천시스템이 이용되고 있으며 다양한 전자상거래에서 채택하여 운영 중에 있다. 추천시스템 중 학문적 연구뿐만 아니라 실제 전자상거래에서 많이 채택하고 있는 방식이 협력적 필터링 접근법에 의한 추천시스템이다. 협력적 필터링 기법은 전자상거래 사이트의 고객들의 혹은 회원들의 상품에 대한 선호도 평가치를 이용하여 특정 고객에 대한 특정 상품의 선호도를 예측할 수 있는 시스템으로 이러한 예측 선호도는 고객이 직접 상품에 대한 정보검색에 소비하여야 하는 정보검색비용을 크게 줄일 수 있다. 또한 특정 전자상거래 사이트의 추천시스템을 이용하는 고객의 수가 증가하고 상품에 대한 선호도 정보가 축적되면 추천시스템의 정확도가 향상되어 기존 이용고객들의 타 전자상거래로의 이탈을 방지할 수 있는 장벽을 마련할 수 있는 등 다양한 이점이 있다. 그러나 추천시스템은 수많은 고객들의 선호 정보를 반영하기 때문에 대부분의 시스템 사용자들과 선호도가 다른 고객들 혹은 부적절한 선호 정보를 제공한 사용자들에게는 정확한 추천을 제공할 수 없을 뿐만 아니라 이러한 특이 성향의 사용자들의 선호 정보로 인하여 추천시스템의 예측 성능이 크게 훼손될 우려가 있다. 본 연구는 대다수 사용자의 선호 정보와 성향이 다른 특이 사용자들을 선호도 예측 이전에 선정하거나 필터링 하기 위한 기준 선정의 필요성에 의해 연구를 진행하였으며 선정 기준을 통해 특이 사용자로 선택된 고객이 실제 시스템에서 선택되지 않은 고객들과 선호도 예측 정확도에 차이가 있는지를 검정하였다.

## 2. 관련 연구

최근 전자상거래에서 정보과잉의 문제를 해결하기 위한 자동화된 기법의 필요성이 커졌으며 추천시스템은 전자상거래에서 고객들 자신의 특정 목적에 부합하는 제품과 정보를 신속하게 찾도록 도와주는 중요한 도구로 받아들여지고 있다(Resnick and Varian, 1997; Schafer et al., 1999). 협력적 필터링 추천시스템의 선호도 예측 성능 향상에 대한 연구는 문헌적으로 많이 이루어져 있지만(Lee, 2006a; Lee et al., 2006; Lee, 2006b; Lee and Lee, 2006) 추천시스템의 선호도 예측에 악영향을 미치는 요소에 관한 주제는 최근에 관심을 보이게 되었다(Lam and Riedl, 2004). O'Mahony 등은 추천시스템의 예측 성능을 저하시키는 요인을 추천시스템의 노이즈로 정의하고 있으며 이러한 노이즈를 다음과 같이 분류하여 정의하고 있다(O'Mahony, 2006).

- 자연적 노이즈(Natural Noise): 자연적 노이즈는 추천시스템이 고객의 선호도를 추정하거나 선호도에 관한 데이터를 수집하는 방법에서 발생할 수 있는 노이즈이다. 일반적으로 추천시스템은 시스템의 고객이 보았거나 구매한 상품에 대한 선호도를 명시적인 평가치, 즉 수치적 척도의 형태로 수집하고 있다. 그러나 모든 인간의 행동은 오류를 범하기 때문에 명시적 선호도를 시스템에 입력하는 과정을 귀찮거나 번거로운 작업으로 여길 수 있다. 그래서 선호도를 시스템에 입력과정에서 고객들은 고의적이 아니더라도 부정확한 데이터를 시스템에 입력할 수 있으며 이러한 형태의 입력 데이터 오류를 자연적 노이즈라고 정의하고 있다. 암시적 자료를 이용하여 고객의 선호도를 추정하는 시스템(예를 들어 웹사이트 로그분석을 통한 선호도 추정)에서도 오류가

발생할 수 있으며 명시적 선호도 평가치 보다 오류가 더 커질 가능성이 있다.

- 고의적 노이즈(Malicious Noise): 자연적 노이즈보다 더 심각한 문제는 시스템에 고의적으로 입력된 노이즈이다. 고의적 노이즈는 예측결과를 의도적으로 유도하기 위해 추천시스템에 고의적인 오류를 입력하는 유형이다. 대부분의 추천시스템은 상업적 목적으로 운영되기 때문에 악의적인 의도를 가진 고객들은 시스템 조작에 대한 강한 동기부여를 받을 수 있다. 예를 들어 책의 저자가 자신의 책의 추천을 늘리기 위해 추천시스템에 인위적인 노력을 가하여 높은 평가치를 부여하거나 혹은 동일 장르의 경쟁 작품에 대한 추천을 줄이기 위해 낮은 평가치를 부여하여 판매를 줄일 수 있다.

고의적인 노이즈는 고객에 대한 추천시스템의 신뢰도를 떨어뜨릴 수 있기 때문에 시스템에 대한 공격(attack)으로도 정의된다. 다음은 시스템에 대한 공격에 대한 분류이다.

- push attack: 이러한 유형의 공격은 시스템에서 좀 더 많은 추천이나 혹은 선호도 예측 결과를 높이기 위해 고의적으로 높은 선호도 평가치를 시스템에 입력하며 상품에 대한 추천을 높여 판매증대를 유도하고자 하는 목적의 공격 형태이다.
- nuke attack: 이러한 유형의 공격은 추천률을 높이거나 선호도 예측의 결과를 높이기 위한 공격과는 달리 반대로 추천률을 하락시키고 예측 결과를 떨어뜨리려 하는 의도로 고의적인 선호도 평가치를 입력하는 형태이다.

협력적 필터링 추천시스템은 이와 같이 외부에서 악의적인 의도에 대해 보안이 약한 측면이 나타났으며 이를 해결하기 위한 연구들이 활발히 진행되고 있다(Lam and Riedl, 2004; O'Mahony, 2006).

## 2.1 추천 알고리즘

최초의 자동화된 협력적 필터링 알고리즘은 GroupLens에서 제시되었으며 추천 대상 고객과 이웃 고객들을 이용하여 선호도를 예측하게 된다(Resnick et al., 1994). 다음은 식(1)은 GroupLens에서 제안한 이웃 기반의 협력적 필터링 알고리즘(Neighborhood Based Collaborative Filtering Algorithm)이다.

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

여기서  $\hat{U}_x$ 는 상품  $x$ 에 대한 추천 대상 고객  $u$ 의 선호도 예측치이며  $\bar{U}$ 는 추천 대상 고객  $u$ 가 평가한 모든 상품에 대한 평균이며  $J_x$ 는 상품  $x$ 에 대한 이웃 고객  $j$ 의 선호도 평가치이고  $\bar{J}$ 는 이웃 고객  $j$ 가 평가한 모든 상품에서 상품  $x$ 에 대한 평가치를 제외한 선호도의 평균이다.  $r_{uj}$ 는 고객  $u$ 와  $j$ 의 선호도 유사 정도를 나타내는 유사도 가중치로 여러 형태로 정의될 수 있으며 피어슨 상관계수가 최초로 사용되었으며 본 연구에서도 피어슨 상관계수를 유사도 가중치로 이용한다.

## 2.2 예측 정확도 평가척도

협력적 필터링에서 예측의 정확도를 평가하기 위하여 일반적으로 시스템의 성능 정확도는 실제 선호도와 필터링을 통한 예측 선호도의 절대평균오차(Mean Absolute Error)를 이용하며 MAE가 크면 전체 시스템의 예측 정확도가 낮아지는 것이고 MAE가 작으면 예측 정확도가 높아진다(Shardanand and Maes, 1995). 다음 식(2)는 MAE의 계산식이다.

$$MAE = \frac{1}{N} \sum_{j=1}^N |R_{u_j} - \widehat{R}_{u_j}| \quad (2)$$

여기서,  $R_{u_j}$ 는 아이템  $j$ 에 대한 고객  $u$ 의 실제 선호도 평가치이고  $\widehat{R}_{u_j}$ 는 아이템  $j$ 에 대한 고객  $u$ 의 선호도 평가치의 예측치이다.

## 3. 연구설계

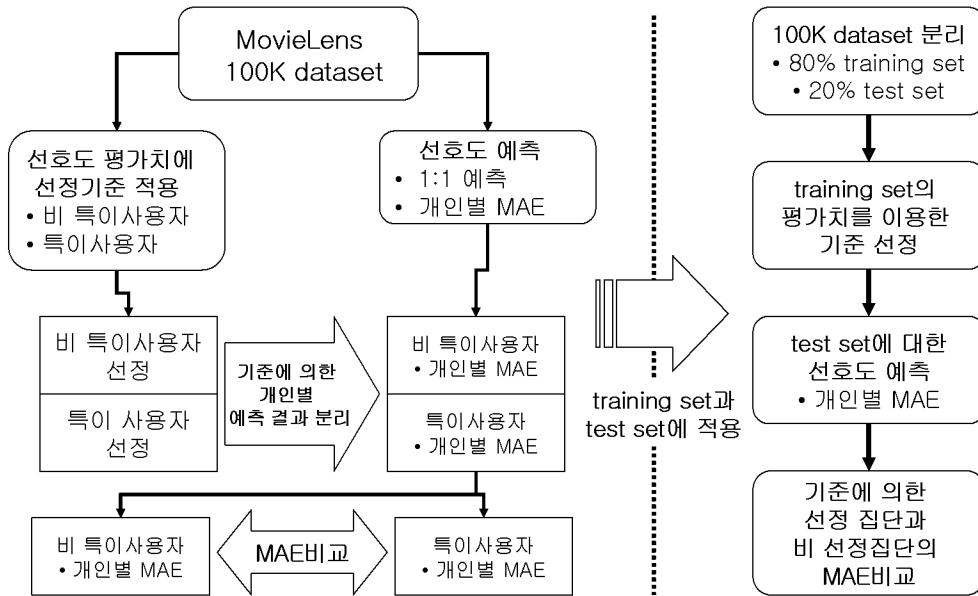
### 3.1 연구 dataset

본 연구는 협력적 추천시스템에서 추천 정확도에 악영향을 미칠 수 있고 예측 정확도가 낮은 고객들을 선호도 예측 이전에 사전 선정 할 수 있는 방법에 대해 연구하였다. 선호도 예측 성과가 낮은 고객을 선정하기 위한 기준을 설정하기 위해 먼저 dataset의 선호도 평가치에 대한 예측을 실시하였다. 일반적으로 선호도 예측 알고리즘의 정확도를 평가하기 위해 훈련집합과 실험집합으로 분리하여 실험집합에 대한 선호도 예측 정확도의 분석 방법이 아닌 전체 data에 대하여 1:1예측을 실시하여 개인별 예측 정확도를 구하였다. 본 연구에서 사용한 dataset은 GroupLens에서 제공하는 MovieLens dataset으로 선정기준의 성능을 평가하기 위하여 100K dataset에 대하여 평가하였다. 100K dataset은 943명의 사용자들이 1682편의 영화에 대해 1-5점 척도를 이용하여 선호도를 평가한 자료로 총 100,000개의 선호도 평가치로 구성되어 있으며 고객에 대한 정보와 영화에 대한 정보가 포함되어 있다.

### 3.2 연구흐름

본 연구는 다음 [그림1]과 같은 과정을 통하여 연구를 진행하였다. 먼저 선호도 예측 성과가 낮은 사용자들을 특이 성향의 사용자로 정의한다. 특이 성향의 사용자 선정기준의 적합성을 평가하기 위하여 100K dataset의 선호도 평가정보를 기준에 따라 비 특이 성향의 사용자와 특이 성향의 사용자 그룹으로 나누었으며 이들의 선호도 예측 성과를 검증하기 위해 전체 dataset에 대한 예측을 실시하여 각 사용자별 개인 MAE를 구하였다. 선정 기준에 따라 선정된 그룹의 개인별 MAE와 선정되지 않은 그룹의 개인별 MAE에 대하여 그룹간의 MAE에 대한 차이를 검증하기 위하여 독립 2표본 t검정을 실시하여 평균차를 검증하였다. 검증 결과를 바탕으로 동일한 선정기준이 80%의 training set과 test set의 선호도 예측에 있어서도 효과가 있는지를 검증하

기 위하여 먼저 20%의 test set에 대한 선호도를 예측하고 개인별 MAE를 구하였다. 선정기준을 80%의 training set에 적용하여 특이 사용자와 비 특이 사용자를 구분하고 두 그룹의 개인별 MAE의 평균에 대해 t검정을 실시하였다.



[그림1]연구 흐름도

### 3.3 특이 사용자 선정기준

특이 사용자와 비 특이 사용자를 구분하기 위한 선정기준은 다음과 같다. 먼저 100K dataset 전체와 80%의 training set에서 사용자의 선호도 평가치  $R_i$ 는 다음과 같이 정의한다.

$$R_i = i, \text{ where } i = 1, 2, 3, 4, 5$$

사용자의 선호도 예측 이전의 사전 정보인 선호도 평가치를 이용하여 다음과 같이 Kronecker delta  $\delta_{u1}, \delta_{u2}, \delta_{u3}$  함수를 다음 식(3)(4)(5)와 같이 정의 한다.

$$\delta_{u1} = \begin{cases} 1, & f_u(R_5) \geq f_u(R_2) \\ 0, & elsewhere \end{cases} \quad (3)$$

식(3)에서  $\delta_{u1}$ 은 시스템의 사용자  $u$ 가 상품에 대해 평가한 선호도 평가치 중  $R_5$ 의 빈도인  $f_u(R_5)$ 가  $R_2$ 의 빈도인  $f_u(R_2)$  보다 크거나 같을 경우 1, 그렇지 않을 경우를 0

으로 정의한다.

$$\delta_{u2} = \begin{cases} 1, & f_u(R_1) \geq f_u(R_4) \\ 0, & \text{elsewhere} \end{cases} \quad (4)$$

식(4)에서  $\delta_{u2}$ 는 사용자  $u$ 의 평가치 중  $R_1$ 의 빈도인  $f_u(R_1)$ 가  $R_4$ 의 빈도인  $f_u(R_4)$ 보다 크거나 같을 경우 1, 그렇지 않을 경우를 0으로 정의한다.

$$\delta_{u3} = \begin{cases} 1, & f_u(\{R_1\} \cup \{R_5\}) \geq f_u(\{R_2\} \cup \{R_3\} \cup \{R_4\}) \\ 0, & \text{elsewhere} \end{cases} \quad (5)$$

식(5)에서  $\delta_{u3}$ 는 사용자  $u$ 의 평가치 중  $R_1, R_5$ 의 빈도의 합인  $f_u(\{R_1\} \cup \{R_5\})$ 가  $R_2, R_3, R_4$ 의 빈도의 합인  $f_u(\{R_2\} \cup \{R_3\} \cup \{R_4\})$ 보다 크거나 같을 경우 1, 그렇지 않을 경우를 0으로 정의한다.

$\delta_{u1}, \delta_{u2}, \delta_{u3}$ 를 이용하여 항등함수  $g(\delta_{u1} \cdot \delta_{u2}), g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 를 각각 다음 식(6)(7)과 같이 정의 한다.

$$g(\delta_{u1} \cdot \delta_{u2}) = \delta_{u1} \cdot \delta_{u2} \quad (6)$$

$$g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3}) = \delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3} \quad (7)$$

#### 4. 분석결과

$g(\delta_{u1} \cdot \delta_{u2}), g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 의 함수 값에 의해 분류된 그룹별 개인 MAE가 차이가 있는지를 독립 2표본 t검정에 의해 분석 하였다. 다음 [표1]은 식(6)의  $g(\delta_{u1} \cdot \delta_{u2})$ 조건에 따라 100K dataset의 전체 자료에서 분류된 두 그룹의 개인 MAE에 대한 독립 2표본 t검정 결과와 80%의 training set과 20%의 test set으로 구분한 dataset에서의 두 그룹의 개인 MAE에 대한 독립 2표본 t검정 결과이다.

[표1]  $g(\delta_{u1} \cdot \delta_{u2})$ 의 조건에 따른 그룹 간 독립 2표본 t검정 결과

dataset	그룹	빈도	평균	t값	유의확률
100K	$g(\delta_{u1} \cdot \delta_{u2}) = 0$	922	0.568	10.07	0.000**
	$g(\delta_{u1} \cdot \delta_{u2}) = 1$	21	0.921		
100K 8:2 dataset	$g(\delta_{u1} \cdot \delta_{u2}) = 0$	921	0.763	4.86	0.000**
	$g(\delta_{u1} \cdot \delta_{u2}) = 1$	22	1.014		

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

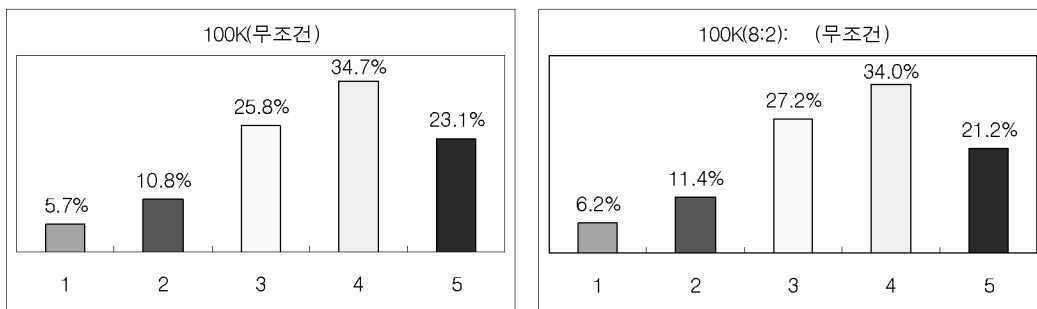
$g(\delta_{u1} \cdot \delta_{u2})=1$  의 조건에서 선정된 특이 사용자들은 100K 전체 dataset에서 943명 중 21명이 선정되었으며 비 특이 사용자들은 922명이 선정되었다. 두 그룹의 MAE의 평균은 각각 0.921과 0.568로 분석되었으며 통계적으로 유의한 평균차가 있음을 알 수 있다. 또한 8:2로 분리한 dataset에서는 특이 사용자가 22명 선정되었으며 비 특이 사용자는 921명으로 선정되었다. 두 그룹의 MAE의 평균은 각각 1.014와 0.763으로 전체 data에 대하여 예측한 MAE의 평균보다 높지만 두 그룹간의 MAE의 평균에 있어 통계적으로 유의한 차이가 있음을 알 수 있다. 다음 [표2]는 식(7)의  $g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 조건에 따라 분류된 두 그룹의 개인 MAE에 대한 독립 2표본 t검정 결과이다.

[표2]  $g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 의 조건에 따른 그룹 간 독립 2표본 t검정 결과

dataset	그룹	빈도	평균	t값	유의확률
100K	$g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})=0$	928	0.569	9.16	0.000**
	$g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})=1$	15	0.951		
100K 8:2 dataset	$g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})=0$	928	0.765	4.01	0.000**
	$g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})=1$	15	1.042		

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

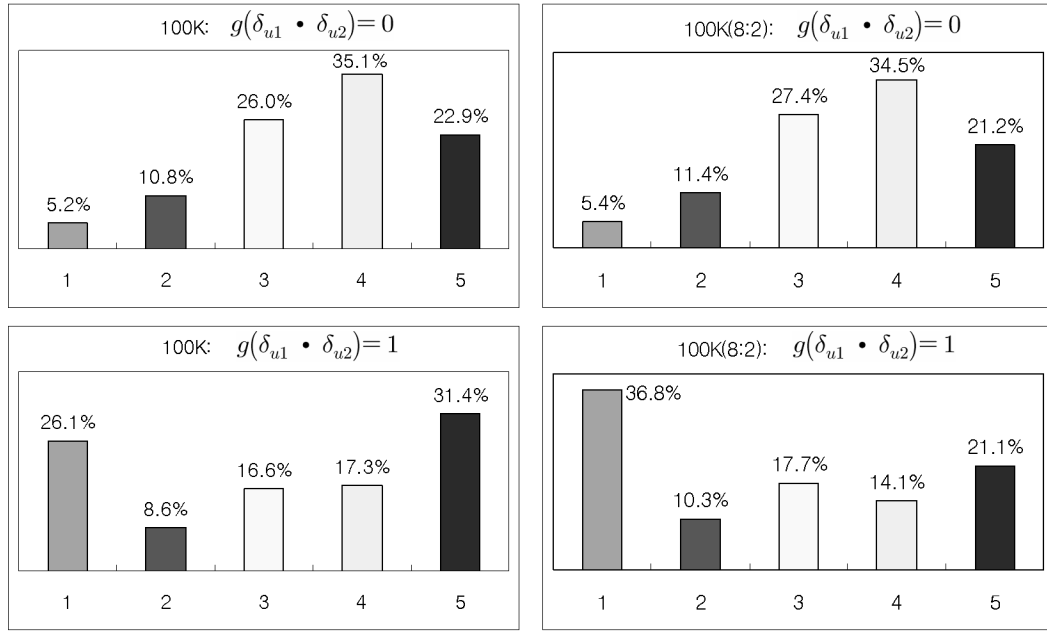
$g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})=1$  의 조건에서 선정된 특이 사용자들은 100K 전체 dataset에서 943명 중 15명이 선정되었으며 8:2로 분리된 dataset에서도 15명이 선정되었다. 100K 전체 data와 8:2로 분리된 dataset 모두에서 선정된 그룹의 평균이 높음을 알 수 있으며 유의수준 0.001에서 통계적으로 유의한 결과로 분석되었다. [표1]의 결과와 비교하여 [표2]에서 선정된 특이 사용자 그룹의 MAE의 평균이 보다 높음을 알 수 있으며  $g(\delta_{u1} \cdot \delta_{u2})=1$ 의 기준에 의해 선정된 특이 사용자들 보다  $g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})=1$ 의 기준에 의해 선정된 특이 사용자 그룹의 선호도 예측 성과가 낮음을 알 수 있다. 다음 [그림2]는 100K 전체 dataset과 8:2로 분리된 dataset의 선호도 평가치의 분포이다.



[그림2] 100K, 8:2 dataset의 선호도 평가치 분포

[그림2]에서 100K dataset과 100K 8:2 dataset의 선호도 평가치는 3을 기준으로 오

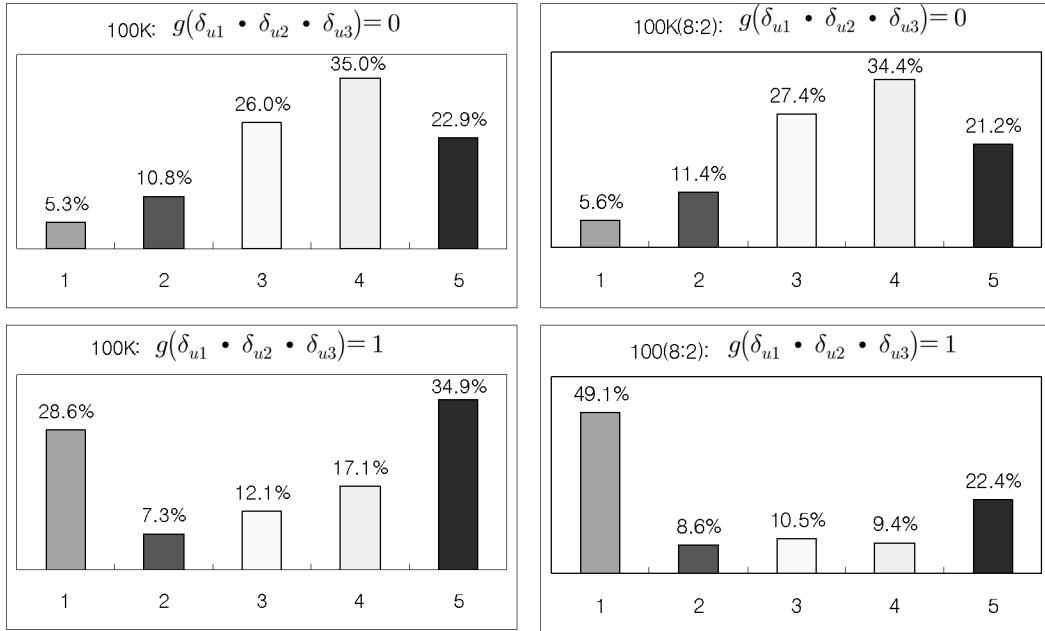
른쪽으로 기울어져 있는 형태를 갖추고 있다. [그림3]은 100K dataset과 100K 8:2 dataset의 선호도 평가치를  $g(\delta_{u_1} \cdot \delta_{u_2})$ 의 선정 기준에 따라 특이 사용자와 비 특이 사용자로 구분하고 그 선호도 평가치의 분포를 나타내고 있다.



[그림3]  $g(\delta_{u_1} \cdot \delta_{u_2})$ 의 선정 기준에 따른 특이 사용자와 비 특이 사용자의 선호도 평가치 분포

[그림3]은  $g(\delta_{u_1} \cdot \delta_{u_2})$ 의 선정 기준에 따라 분류된 특이 사용자와 비 특이 사용자의 선호도 평가치의 분포로 비 특이자로 선정된 사용자들의 선호도 평가치는 [그림2]의 무조건일 경우의 선호도 평가치와 유사한 유형을 보이며 특이 사용자의 경우 선호도 평가치 1과 5의 비율이 높은 쌍봉형의 분포를 나타내고 있음을 알 수 있다. 100K 8:2 dataset의 분포에서는 선호도 평가치 1의 분포가 5에 비하여 높음을 알 수 있다. [그림4]는  $g(\delta_{u_1} \cdot \delta_{u_2} \cdot \delta_{u_3})$ 의 선정 기준에 따라 분류된 특이 사용자와 비 특이 사용자의 선호도 평가치 분포이다.





[그림4]  $g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 의 선정 기준에 따른 특이 사용자와 비 특이 사용자의 선호도 평가치 분포

[그림4]는  $g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 의 선정 기준에 따라 분류된 특이 사용자와 비 특이 사용자의 선호도 평가치 분포로 비 특이 사용자로 선정된 사용자들의 선호도 평가치는  $g(\delta_{u1} \cdot \delta_{u2})$ 의 선정 조건의 비 특이 사용자들의 분포 유형과 유사하고 또한 전체 평가치의 분포와 유사한 것을 알 수 있다. 특이 사용자들의 선호도 평가치 분포는  $g(\delta_{u1} \cdot \delta_{u2})$ 의 선정 기준에 따라 분류된 특이 사용자들의 분포와 유사하지만 100K 전체 dataset에서는 선호도 평가치 중간값의 비율이 더 낮은 특징이 있다.

## 5. 결론

본 연구는 추천시스템에서 선호도 예측 정확도가 낮은 사용자들을 선호도 예측 이전에 분류하기 위한 기준에 대하여 연구하였다. 본 연구의 결과에서 특이 사용자를 구분하기 위해  $g(\delta_{u1} \cdot \delta_{u2})$ 와  $g(\delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3})$ 의 기준에 따라 분류된 두 그룹의 사용자들은 선호도 예측 정확도에서 통계적으로 유의한 차이가 있음을 알 수 있으며 이는 사용자들의 선호도 평가치 자료만을 이용하여 선호도 예측 이전에 사전 필터링의 기준으로 활용할 수 있을 것으로 기대된다. 본 연구를 통하여 전체 dataset의 선호도 평가치 분포의 특성을 이용한 사전 필터링 방법은 8:2로 분할한 dataset에서도 유용한 필터링 방법임을 알 수 있으며 이는 실제 추천시스템에서도 dataset의 선호도 평가치의 분포 특성을 이용하여 사전 필터링을 통한 노이즈 제거가 가능할 것으로 판단된다. 또한 분류된 특이 성향의 사용자들을 분석함으로써 특이 성향의 사용자들이 실제 선호도 예측에 있어 어떠한 영향을 미치고 노이즈로써 제거되어야 할 것인지 그렇지

않으면 특히 성향의 사용자에게 대하여 가중치를 주어야 할 것인지에 대한 판단 기준으로써 이용될 수 있을 것이다. 차기 연구로는 사전 필터링 기준을 적용하여 협력적 필터링 알고리즘의 선호도 예측 성능을 향상시키는 방법과 이를 통한 Top-N 추천의 정확도를 향상시키는 방법에 대한 연구가 필요하다.

### 참고문헌

1. Lam, S. K. and Riedl, J. (2004). Shilling recommender systems for fun and profit, *In Proceedings of the 13th international conference on World Wide Web*, New York, NY: ACM Press, 393-402.
2. Lee, H. C. (2006a). Improved Algorithm for User Based Recommender System, *Journal of the Korean Data & Information Society*, 17(3) 717-726.
3. Lee, H. C., Lee, S. J. and Chung, Y. J. (2006). The Effect of Co-rating on the Recommender System of User Base, *Journal of the Korean Data & Information Society*, 17(3), 775-784.
4. Lee, H. C. (2006b). On the Effect of Significance of Correlation Coefficient for Recommender System, *Journal of the Korean Data & Information Science Society*, 17(4), 1129-1139.
5. Lee, H. C. and Lee, S. J. (2006). On the Study of Perfect Coverage for Recommender System, *Journal of the Korean Data & Information Society*, 17(4), 1151-1160.
6. O'Mahony, M. P., Hurley, N. J., Gu, nol and Silvestre, C. M. (2006). Detecting noise in recommender system databases, *In Proceedings of the 11th international conference on Intelligent user interfaces*, Sydney: ACM Press, 109-115.
7. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *In Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'94)*, October 22-26, 175-186.
8. Resnick P. and Varian. H. R. (1997). Recommender systems, introduction to the special section. *Communications of the ACM*, 40(3), 56-58.
9. Shardanand, U. and Maes, P. (1995). Social information filtering: algorithms for automating 'word of mouth', *In Proceedings of the SIGCHI conference on Human factors in computing systems*, Denver, Colorado: ACM Press, 210-217.

[ 2007년 6월 접수, 2007년 7월 채택 ]