

A Study on the Interrelationship between the Prediction Error and the Rating's Pattern in Collaborative Filtering

Seok-Jun Lee¹⁾ · Sun-Ok Kim²⁾ · Hee-Choon Lee³⁾

Abstract

Collaborative filtering approach for recommender systems are now widely applied in e-commerce to assist customers to find their needs from many that are frequently available. this approach makes recommendations for users based on the opinions to similar users in the system. But this approach is opened to users who present their preference to items or acquire the preference information form other users, noise in the system makes significant problem for accurate recommendation. In this paper, we analyze the relationship between the standard deviation of preference ratings for each user and the estimated ratings of them. The result shows that the possibility of the pre-filtering condition which detecting the factor of bad effect on the prediction of user's preference. It is expected that using this result will reduce the possibility of bad effect on recommender systems.

Keywords : 고객의 사전정보, 사전평가, 협력적 필터링

1. 서 론

인터넷은 기업들이 사업을 하기 위한 수단으로 급격히 성장하고 있다. 기업들은 인터넷을 통하여 쉽게 광고를 할 수 있으며 쉽게 제품과 서비스를 판매할 수 있고 고객들과의 의사소통 또한 원활히 할 수 있다. 인터넷을 통한 마케팅과 광고는 기업의 큰 기회를 약속하고 있다. 또한 고객의 입장에서 인터넷을 통한 구매는 과거의 거래과

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

정에 소요되는 시간을 단축시키고 있으며 손수 제품에 대한 정보를 얻기 위하여 주변에서 자문을 구하고 제품 카탈로그 등에서 정보를 탐색하던 시간과 노력을 줄이고 있다. 인터넷을 통한 전자상거래는 고객에 대한 잠재적 서비스를 제공할 수 있는 수많은 장점들을 가지고 있으며 잠재적 서비스로는 고객의 편의, 제품에 대한 폭넓은 선택, 더 나은 제품 정보, 새로운 형태의 제품이나 서비스, 그리고 가격비교를 통한 제품 선택 등을 들 수 있다. 그러나 고객에 대한 잠재적 서비스의 유형이 증가하면 할수록 제품과 서비스에 대한 더 많은 정보를 생성하게 되고 과거의 제품 선택과 거래 과정의 시간을 단축시키고는 있지만 새로운 정보의 생성을 통한 정보검색의 비용을 높이고 있다. 이러한 정보검색의 비용은 제품선택에 대한 고객의 욕구를 떨어뜨릴 수 있기 때문에 이를 해결하기 위한 노력들이 진행되고 있고 또한 수많은 정보들 중 개별 고객에게 필요한 정보를 제공할 수 있는 개인화 서비스가 필요하게 되었다.

2. 연구목적

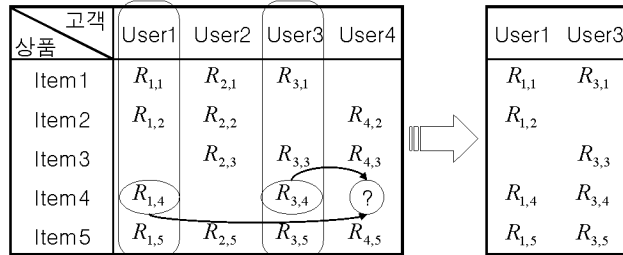
본 연구는 협력적 필터링 기법의 선호도 예측 알고리즘의 예측 정확도와 선호도 예측 이전에 주어진 고객의 사전 정보인 선호도 평가치의 통계적 특성과의 관계에 대하여 연구하였다. 기존의 연구에서는 선호도 예측 알고리즘의 정확도를 높이기 위하여 선호도 예측 알고리즘을 조정하고 이웃고객과의 선호도 유사성을 나타내는 유사도 가중치를 조정하여 이에 대한 효과를 분석하는 연구가 많이 진행되었으며 신속하고 보다 정확한 예측을 위하여 선정된 이웃에서 선호도 유사성이 높은 고객들을 선정하는 이웃선정의 기법에 대하여 많은 연구가 진행되었다. 그러나 본 연구에서는 선호도 예측 알고리즘이나 유사도 가중치 혹은 이웃선정의 기법을 통한 선호도 예측 정확도 향상에 관한 관점보다 선호도 예측 이전에 고객들이 평가한 선호도 평가치의 통계적 특성을 분석하여 선호도 예측 이전에 예측의 정확도를 사전에 평가가 가능한지에 대한 관련성을 분석하였다. 이는 추천시스템에서 선호도 예측 이전에 고객의 사전정보를 이용하여 선호도 예측의 성과가 낮은 고객들을 선정하여 그들이 선호도 예측에 미치는 영향을 선호도 예측에서 제거하거나 혹은 선호도 예측 과정에서 보정하여 예측 결과를 향상시킬 수 있는 가능성을 제시한다. 또한 추천시스템에서 선호도 예측의 정확도가 낮은 고객들의 특성을 파악할 수 있는 가능성을 제시하는 것을 이 논문의 목적으로 한다.

3. 추천시스템

추천시스템은 제품, 서비스, 정보(일반적으로 통칭하여 상품) 중 고객에게 가장 적합한 상품을 추천하기 위해 다양한 응용분야에 이용되고 있다. 특히 전자상거래에서 추천시스템은 인터넷과 결합하여 고객과 상품 간의 관계(구매, 선호도, 그리고 카탈로그 검색 등과 같은 활동)에서 얻어지는 거래 데이터와 인터넷상에서 실시간으로 얻어지는 고객 행동 데이터를 분석하여 얻어진 결과를 이용하여 상품 추천을 위한 핵심적인 정보를 수집한다. 추천시스템은 다양한 접근법이 연구되어 왔으며 실제 전자상거래에서는 협력적 필터링 기법을 이용한 시스템들이 일반적으로 적용되고 있다.

3.1 협력적 필터링

협력적 필터링은 추천시스템 연구의 근간을 이루고 있다. 또한 실제 많은 전자상거래 추천시스템에 적용되어 사용되고 있다(Schafer 등, 1999). 최초의 자동화된 협력적 선호도 예측 알고리즘은 GroupLens에 의해 제안되었으며 추천 대상 고객과 관련성이 있는 고객들을 이웃으로 선정하여 상품들에 대한 두 고객의 선호도를 고려한 이웃 기반의 협력적 필터링 알고리즘(Neighborhood Based Collaborative Filtering Algorithm)을 적용하여 유즈넷 뉴스(UseNet News) 그룹의 기사를 추천하였다(Resnick 등, 1994). 협력적 필터링에서 일반적으로 상품과 고객간의 관계는 매트릭스 구조로 이루어져 있으며 상품에 대한 고객의 선호도 평가자료 혹은 구매자료의 형태로 정의된다. 이때 선호도 예측을 위한 알고리즘의 적용은 먼저 추천 대상 고객과 이웃 고객의 선정에서 시작된다.



<그림 1> 추천대상고객의 이웃선정 과정

<그림 1>에서 고객 User4의 상품 Item4에 대한 선호도 예측을 위해 선정된 이웃은 해당 상품에 대하여 선호도 혹은 구매 정보(여기서는 선호도 평가치)를 가지고 있는 고객 User1과 User3이 선택된다. 협력적 필터링에서 고객의 이웃 선정과정은 추천대상 상품에 선호도를 표기한 이웃이 선택되기 때문에 선택된 이웃에 영향을 받게 되며 이때 선택된 이웃이 불성실한 응답이나 악의적인 평가를 하였을 경우 시스템에 노이즈가 될 가능성이 매우 크다(Lam, Riedl 2004). 이러한 노이즈를 찾기 위한 다양한 방법들이 제시되고 있다(Chirita 등, 2005; O'Mahony 등, 2006).

3.2 유사도 가중치

선정된 이웃과 추천 대상고객과의 상품에 대한 유사정도는 유사도 가중치로 정의한다. 유사도 가중치는 두 고객, 즉, User1과 User4 혹은 User3과 User4의 관계와 같이 두 고객 간의 선호관계를 정의하며 피어슨 상관계수, 벡터 유사도 등이 이용되며 일반적으로 피어슨 상관계수를 많이 이용한다. 다음은 두 고객의 유사도 가중치로 정의된 피어슨 상관계수이다.

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

3.3 선호도 예측 알고리즘

상품에 대한 고객의 선호정도를 나타내는 선호도 평가치를 이용한 선호도 예측을 위해 GroupLens에서는 이웃기반의 협력적 필터링 알고리즘(Neighborhood Based Collaborative Filtering Algorithm)을 제안하였으며(Resnick 등, 1994), Lee(2006)의 연구에서는 NBCFA를 개선한 대응평균 알고리즘(Correspondence Mean Algorithm)을 제안하였다.

3.3.1 Neighborhood Based Collaborative Filtering Algorithm

초기 GroupLens 시스템에서는 유즈넷 뉴스 그룹의 모든 고객들의 선호 기사에 대한 상관관계를 정의하였다. 특정 문서를 추천 받을 고객이 해당 기사에 대하여 어느 정도 선호도를 가지는지를 예측하기 위해 다음과 같은 이웃기반의 협력적 필터링 알고리즘(NBCFA)을 제안하였다. NBCFA는 다음과 같이 정의한다.

$$\hat{U}_x = \bar{U} + \frac{\sum_{j \in \text{Raters}} (J_x - \bar{J})r_{uj}}{\sum_{j \in \text{Raters}} |r_{uj}|}, \quad \text{where} \quad \bar{J} = \frac{\sum_{i=1}^n J_i}{n}, \quad i \neq x \quad (2)$$

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 여기서는 피어슨 상관계수를 유사도 가중치로 이용한다.

3.3.2 Correspondence Mean Algorithm

Lee(2006)의 연구에서는 GroupLens에서 제안한 NBCFA를 개선한 대응평균 알고리즘(Correspondence Mean Algorithm)을 제안하여 추천 대상 고객 u 의 선호도 예측 정확도를 향상하였다(이희춘, 이석준, 2006 a; 이희춘, 이석준, 2006 b; Lee, 2006; Lee 등, 2006). CMA는 다음과 같이 정의한다.

$$\hat{U}_x = \bar{U}_{match} + \frac{\sum_{j \in Raters} (J_x - \bar{J}_{match}) r_{uj}}{\sum_{j \in Raters} |r_{uj}|}, \text{ where } \bar{J} = \frac{\sum_{i=1}^n j_i}{n}, i \neq x \quad (3)$$

여기서 \bar{U}_{match} 는 추천 대상 고객 u 와 각 이웃 고객 j 가 공통으로 평가한 상품들의 평가치의 평균들 즉, 상관계수를 계산할 때 이용되는 평가치들의 평균을 구하고 이 평균들을 다시 평균을 계산한다. 이는 NBCFA에서 과도하게 반영된 추천 대상 고객 u 의 선호도를 조정하는 역할을 한다. NBCFA에서 사용되는 \bar{J} 는 이웃 고객 j 의 전체 선호도 평가치의 평균이 되는 반면 CMA의 \bar{J}_{match} 는 추천 대상 고객 u 와 이웃 고객 j 가 공통으로 평가한 상품들에 대한 선호도 평가치만을 이용한 평균이고 또한 이는 NBCFA에서 이웃 고객 j 의 선호도가 과도하게 적용되어 있는 것을 조정하는 역할을 한다.

3.3.3 선호도 예측 정확도 평가척도

협력적 필터링에서 예측의 정확도를 평가하기 위하여 일반적으로 시스템의 성능 정확도는 실제 선호도와 알고리즘을 통하여 생성된 선호도 예측치와의 절대평균오차 (Mean Absolute Error : MAE)를 이용하며 MAE가 크면 전체 시스템의 예측 정확도가 낮아지는 것이고 MAE가 작으면 예측 정확도가 높아진다(Breese 등, 1998). MAE는 다음과 같이 정의한다.

$$MAE = \frac{1}{N} \sum_{j=1}^N |R_{uj} - \hat{R}_{uj}| \quad (4)$$

여기서, R_{uj} 는 상품 j 에 대하여 고객 u 가 평가한 실제 선호도 평가치이며, \hat{R}_{uj} 는 선호도 예측 알고리즘으로 예측된 상품 j 에 대하여 고객 u 의 선호도 예측치이다.

4. 실험설계 및 결과

4.1 실험설계

본 연구는 협력적 필터링 선호도 예측 알고리즘의 예측 성능 평가척도인 MAE와 선호도 예측 이전의 고객에 대한 사전정보인 선호도 평가치의 통계량과의 관련성을 분석하였다. 분석을 위하여 고객들이 평가한 선호도 평가치들의 통계량들을 NBCFA와 CMA에 의한 예측결과의 MAE와 관련성을 조사하였다. 선호도 예측 정확도는 MovieLens 100K dataset을 이용하였으며 NBCFA와 CMA를 통한 결과를 산출하였다. 예측 알고리즘을 통한 예측의 정확도를 높이고 예측 결과와 사전정보인 선호도 평가치와의 통계량과의 관련성을 높이기 위하여 dataset의 모든 선호도 평가치를 예측하였으며 결과를 바탕으로 80%의 training set과 20%의 test set에도 적용시켜 실제 예측 시스템에 대한 적용 가능성에 대하여 실험하였다. 일반적으로 MAE는 전체

시스템의 정확도를 평가하기 위하여 사용되기 때문에 개별고객의 선호도 평가치 특성과 예측 오차와의 관계를 조사하기 위하여 선호도 예측 결과를 개별 고객별로 분류하여 개별 고객의 실제 평가치와 선호도 예측치 간의 MAE인 개인별 MAE를 계산하였다.

4.2 실험결과

4.2.1 예측 알고리즘의 MAE와 사전정보의 통계량과의 상관성

선호도 예측 알고리즘인 NBCFA와 CMA에 의해 생성된 예측치를 개인별 실제 평가치와 비교하여 개인별 MAE를 구하고 선호도 예측의 사전 정보인 선호도 평가치의 통계량을 구하여 관련성을 파악하였다. 다음 <표 1>은 선호도 평가치의 통계량인 평균, 표준편차, 분산, 변동계수, 첨도, 왜도, 범위, 사분범위와 개인별 MAE의 상관관계를 파악하기 위한 산점도이다.

선호도 예측 오차인 개인별 MAE와 통계량과의 산점도를 조사한 결과 표준편차와 변동계수(coefficient of variation)의 관련성이 타 통계량에 비하여 관련성이 높은 것으로 나타났다.

다음 <표 2>는 변동계수와 MAE의 상관계수이다.

<표 2> NBCFA와 CMA의 개인별 MAE와 변동계수의 상관계수

알고리즘	상관계수
NBCFA의 MAE	0.7984
CMA의 MAE	0.7163

분석결과 NBCFA와 CMA의 선호도 예측 오차인 개인별 MAE와 선호도 평가치의 변동계수와는 매우 높은 상관관계를 나타냄을 알 수 있으며 NBCFA에 의한 MAE와의 상관계수는 0.7984로 나타났으며 CMA의 경우 0.7163으로 조금 낮은 상관계수를 나타냄을 알 수 있다.

다음 <표 3>은 표준편차와 MAE의 상관관계를 보여주는 산점도와 상관계수이다.

<표 3> NBCFA와 CMA의 개인별 MAE와 표준편차와의 상관계수

알고리즘	상관계수
NBCFA의 MAE	0.8824
CMA의 MAE	0.7929

분석결과 NBCFA와 CMA의 선호도 예측 오차인 개인별 MAE와 선호도 평가치의 표준편차와는 매우 높은 상관관계를 나타냄을 알 수 있으며 NBCFA에 의한 MAE와의 상관계수는 0.8824로 나타났으며 CMA의 경우 0.7929로 조금 낮은 상관계수를 나타냄을 알 수 있다. 두 통계량과의 상관계수 분석을 통하여 변동계수, 표준편차의 상관성이 타 통계량과 비교하여 관련성이 높은 것으로 <표 1>에서 나타났다. 또한 변동계수와 MAE의 관계보다 표준편차와 MAE의 상관관계가 더 높은 것으로 나타났다. 따라서 예측 이전에 개인별 선호도 평가치의 표준편차를 사전에 알고 있다면 개인의

MAE를 대략적으로 예측 할 수 있다는 의미가 되므로 고객의 사전정보를 이용하여 선호도 예측 과정에서 선호도 예측 정확도가 낮은 고객들을 선호도 예측 이전에 선정 할 수 있으며 또한 선정된 고객들의 특성을 사전에 분석 할 수 있다.

<표 1> 선호도 평가치의 통계량과 MAE의 산점도

알고리즘	평균과 MAE 산점도	표준편차와 MAE 산점도	변동계수와 MAE 산점도
NBCFA			
CMA			
	첨도와 MAE 산점도	왜도와 MAE 산점도	사분범위와 MAE 산점도
NBCFA			
CMA			

4.2.2 표준편차의 크기에 따른 MAE의 일원분산분석

분석을 통하여 MovieLens 100K dataset의 전체 자료를 이용한 NBCFA와 CMA의 예측 오차인 개인별 MAE와 선호도 예측 이전의 개별 고객의 사전정보인 선호도 평가치의 표준편차가 관련성이 큼을 알 수 있다. 본 연구에서는 이들의 연관성을 구간별로 구분하여 차이를 검증하고 이를 바탕으로 선호도 예측 이전에 예측 오차가 클 것으로 예상되는 고객들을 잘 구분할 수 있는지에 대하여 알아보았다. 분석을 위하여 개별 고객의 선호도 평가치의 표준편차를 집단을 다음과 같이 3개의 구간으로 나누어 분석하였다. 선호도 평가치 표준편차의 3개구간은 다음과 같이 나누었다. 1구간은 $[\min_s, \bar{x}_s - S_s]$, 2구간은 $[\bar{x}_s - S_s, \bar{x}_s + S_s]$, 3구간은 $[\bar{x}_s + S_s, \max_s]$ 로 각각 나누어 일원분산분석을 하였다. 일원분산분석 결과는 <표 4>와 같다.

<표 4> 일원분산분석(전체자료)

알고리즘	1집단	2집단	3집단	F값	유의확률	다중범위 검정
NBCF의 MAE	.354	.558	.808	667.832	0.000**	{1}{2}{3}
CMA의 MAE	.314	.507	.713	424.046	0.000**	{1}{2}{3}

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

MovieLens 100K dataset 전체 자료에 대한 선호도 평가치 예측 결과를 이용하여 개인별 MAE와 개인별 선호도 평가치의 표준편차와의 관계를 분석한 일원분산분석결과 표준편차의 3개 구간 집단에 따른 NBCFA와 CMA의 MAE는 유의적인 차이를 보였다. 일원분산분석결과에 대해 사후검정은 Duncan, Scheffe의 다중범위검정을 사용하였다. 다중범위검정결과 Duncan, Scheffe의 다중범위검정결과는 동일하게 3개의 부집단으로 나뉘어 나타났으며 표준편차가 작을수록 NBCFA와 CMA의 MAE가 작은 것으로 나타났다. 이는 선호도 예측 이전에 사전정보인 개인별 선호도 평가치의 통계량인 표준편차를 이용하면 선호도 예측 오차가 큰 고객들을 사전에 분류할 수 있음을 의미한다. 전체 자료에 대한 분석결과를 바탕으로 실제 예측 시스템에 적용할 수 있는지를 평가하기 위하여 MovieLens 100K dataset을 training set과 test set으로 분할하여 결과를 다시 확인하였다.

전체 자료에 대한 분석 결과를 바탕으로 80%의 training set과 20%의 test set에서도 유사한 결과를 나타내는지 분석하였다. 먼저 80%의 training set의 개별 사용자들의 선호도 평가치의 표준편차를 계산하고 표준편차에 따라 전체 자료 분석에 적용하였던 방법을 적용하여 3개의 집단으로 구분하였다. training set을 이용하여 20%의 test set의 선호도 평가치에 대하여 NBCFA와 CMA를 적용하여 예측하고 개인별 MAE를 구하여 결과를 비교하였다. 다음 표5는 분할 dataset에 대한 분석결과이다.

<표 5> 일원분산분석(80% training, 20% test set 구분)

알고리즘	1집단	2집단	3집단	F값	유의확률	다중범위 검정
NBCF의 MAE	.609	0.743	1.020	127.964	0.000**	{1}{2}{3}
CMA의 MAE	.599	0.731	0.998	121.851	0.000**	{1}{2}{3}

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

80%의 training set과 20%의 test set 분석결과에 대한 일원분산분석결과 표준편차의 크기에 따라 구분된 3개의 집단의 결과에서도 NBCFA와 CMA에 의한 결과 모두에서 전체 자료에 대한 분석결과와 같이 유의적인 차이가 있음을 알 수 있다. 일원분산분석결과에 대해 사후검정 결과에서도 Duncan, Scheffe의 다중범위검정결과 동일하게 3개의 부집단으로 나뉘었으며 표준편차가 작을수록 NBCFA와 CMA의 MAE가 작은 것으로 나타났다. 따라서 실제 추천시스템의 선호도 예측 과정 이전에 고객의 사전정보를 이용하여 고객들의 예측 오차의 특성을 파악할 수 있음을 알 수 있고 이를 통한 예측 오차가 클 것으로 예상되는 고객들을 선호도 예측 이전에 선정하여 그들의 특성을 파악할 수 있는 가능성을 제시하였다. 또한 기존에 사용되던 선호도 예측 알고리즘인 식(2)와 식(3)을 본 논문의 연구결과를 이용하여 다음과 같은 식을 제안할 수 있다.

$$\hat{U}_x + f(SD) \tag{5}$$

여기서, \hat{U}_x 는 기존의 선호도 예측 알고리즘이며, $f(SD)$ 는 선호도 평가치의 표준편차를 이용한 보정 함수이다. 차후 연구로 식(5)와 같은 보정 알고리즘을 선호도 예측에 적용하여 예측의 정확도를 높이는 연구에 활용 할 수 있을 것으로 기대 된다.

5. 결론

본 연구는 기존에 선호도 예측 알고리즘의 정확도를 높이기 위한 예측 알고리즘에 대한 연구와 달리 선호도 예측 이전에 고객들의 선호도 평가치 자료를 이용하여 예측 정확도에 대한 사전평가 가능성에 대하여 연구하였다. 본 연구를 통하여 개인별 선호도 평가치의 표준편차가 개인별 MAE와 상관성이 높게 나타나는 것을 알 수 있었으며 이는 추천시스템에서 개인별 응답의 표준편차를 사전에 알고 있으면 개인별 MAE의 크기를 사전에 감지 할 수 있음을 의미한다. 또한 연구의 결과를 통하여 개인별 선호도 평가치의 표준편차를 이용하여 선호도 예측 이전에 선호도 예측에 악영향을 줄 수 있는 고객들을 분류할 수 있는 기준을 제시할 수 있을 것이며 선호도 예측 알고리즘에 선호도 평가치의 표준편차를 이용한 보정 알고리즘을 개발하여 선호도 예측의 정확도를 높일 수 있을 것이다. 또한 사전정보에서 얻어진 표준편차를 이용하여 선호도 예측치의 MAE가 클 경우 실제 고객에 대한 상품의 추천방법인 Top-N 추천에서 추천되는 상품의 수를 적정 수준으로 정하는데 기여할 수 가 있다. 향후 연구로 개인별 선호도 평가치의 표준편차가 큰 고객이 다른 이웃에 어떤 영향을 미치는지에 대한 연구와 이들이 예측 시스템에 악영향을 미치는 고객일 경우 추천시스템의 성능

을 인위적으로 조작하려는 공격자인지를 가려내는 연구의 기준을 제시할 수 있을 것으로 기대된다.

참고문헌

1. 이희춘, 이석준 (2006 a). 사용자 기반 추천시스템에서 근접이웃 알고리즘과 수정알고리즘의 예측 정확도에 관한 연구, *Journal of the Korean Data Analysis Society*, Vol.8(5), pp.1893-1904.
2. 이희춘, 이석준 (2006 b). 대응평균 알고리즘을 이용한 협력적 필터링 추천시스템의 성능향상, *한국경영정보학회 추계컨퍼런스*, 서울 한국과학기술회관, pp.208-214.
3. Breese, J. S., D. Heckerman and C. Kadie (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering, *In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence*, 43-52.
4. Chirita, P. A., Nejdil, W. and Zamfir, C.(2005). Preventing shilling attacks in online recommender systems, *In Proceedings of the 7th annual ACM international workshop on Web information and data management*, Bremen, Germany, 67-74.
5. 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, USA, 393-402.
6. Lee, H. C.(2006). Improved Algorithm for User Based Recommender System, *Journal of the Korean Data & Information Society*, Vol.17(3), 717-726.
7. Lee, H. C., Lee, S. J., Chung, Y. J.(2006). The Effect of Co-rating on the Recommender System of User Base, *Journal of the Korean Data & Information Society*, Vol.17(3), 775-784.
8. O'Mahony, M. P., Hurley, N. J., Silvestret, C.M. (2006). Detecting noise in recommender system databases, *In Proceedings of the 11th international conference on Intelligent user interfaces*, Sydney, Australia, 109-115.
9. 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.
10. Schafer, J. B., Konstan, J. and Riedl, J.(1999). Recommender systems in e-commerce, *In Proceedings of the 1st ACM Conference on Electronic Commerce*, November, 158-166.

[2007년 6월 접수, 2007년 8월 채택]