

## An Exploratory Study for Decreasing Error of Prediction Value of Recommended System on User Based<sup>1)</sup>

Hee Choon Lee<sup>2)</sup>

### Abstract

This study is to investigate the error of prediction value with related variables from the recommended system and to examine the error of prediction value with related variables. To decrease the error on the collaborative recommended system on user based, this research explored the effects on the prediction related response pair between raters' demographic variables and Pearson's coefficient and sparsity. The result shows comparative analysis between existing error of prediction value and conditioned one.

**Keywords** : 추천시스템, Correlation, MAE

### 1. 서론

정보시스템의 발전과 웹의 비약적인 발전으로 전자상거래가 주요한 거래수단으로 대두되고 있으며 정보기술의 발전으로 미래의 사회 환경에도 전자상거래는 더욱 발전하며 그 활용성은 증대되리라고 예상된다. 이에 따라 전자상거래 시스템에서 사용되는 추천시스템 또한 적용분야와 활용성은 또한 증대되리라고 생각된다. 이미 추천시스템은 Amazon.com, CDnow.com 등의 유명한 상거래 사이트에서 성공적인 결과를 보이고 있다. 이 논문에서는 추천시스템에서 예측의 오차가 어떤 요인과 관련이 있는지를 탐색하고 오차와 관련이 있는 요인에 대해 예측 오차의 증감에 대해 조사 분석한다. 또한 이 논문에서는 사용자 기반(user based collaborative) 추천시스템에서 오차를 감소시키기 위해 사용자가 가지고 있는 정보와 사용자 기반 추천시스템에서 사용되는 피어슨 상관계수와 응답 쌍은 예측에 어느 정도의 영향을 가지고 있는지를 탐색하였다. 피어슨 상관계수를 근간으로 하는 방식은 여러 분야에서 좋은 결과를 나타냈

---

1) 이 논문은 2003년도 상지대학교 교내 연구비 지원에 의한 것임

2) 강원도 원주시 우산동 660번지 상지대학교 컴퓨터데이터정보학과 교수  
E-mail : choolee@sangji.ac.kr

으나, 가장 큰 문제점은 희소성(sparsity)으로 인해서 발생하는 예측의 정확도 저하이다. 협동적 필터링(collaborative Filtering) 기법은 초기 평가의 문제점, 희소성(sparsity), 모호 집단(gray sheep) 등과 같은 문제점 등이 있다. 초기 평가의 문제점은 시스템 구축 초기에서 사용자로부터 충분한 평가 정보를 받지 못한 경우 정확한 추천이 어려운 문제이며 희소성(sparsity)의 문제는 초기 시스템에서 발생하는 문제와 유사하다. 상품에 대해서 충분한 평가 결과를 구성하기 어려워 사용자와 평가 정보로 구성된 행렬이 희소성(sparsity)을 가지는 문제이다. 모호 집단의 문제는 좋은 것과 싫은 것이 분명하지 않거나 평가 결과가 일정하지 않은 특이한 사용자들은 추천 예측에서 오차를 증가시키므로 추천이 어려운 문제이다. 본 논문에서 Grouplens의 movielens dataset에 대해 탐색적인 분석을 하고 예측오차의 요인에 대해 조사 분석하고자 한다.

## 2. 추천시스템의 분류

김영설(2002)은 일반적으로 추천시스템의 추천방법을 다음과 같이 분류하였다.

- ① 단순검색(Raw retrieval)은 고객이 찾고자 하는 내용에 대한 검색 결과를 제공한다.
- ② 수동선택(Manual selection)은 고객이 추천을 요구하면 해당 분야의 전문가가 고객이 좋아할 제품을 예측하여 제공한다.
- ③ 통계적 요약(Statistical summaries)은 고객이 추천을 요구하면 고객이 속하는 공동체에서 가장 인기있는 제품을 추천한다.
- ④ 속성기반(Attributed-based)은 아이템의 속성에 대한 고객의 선호도에 따라 추천을 수행한다.
- ⑤ 아이템간 관계(Item-to-item correlation)는 상품을 작은 집합으로 구분하여 고객이 특정 집합에 포함된 상품을 구입하면 같은 집합에 포함된 상품들을 고객에게 추천하는 방식이다.
- ⑥ 사용자간 관계(User-to-user correlation)는 추천을 요구하는 고객과 제품을 구매한 다른 고객과의 관계에 기초하여 고객에게 제품을 추천한다. 이 기술은 개인화된 추천을 제공하기 위하여 유사 사용자(그룹)의 의견을 사용하므로 협동적 필터링의 한 유형으로 볼 수 있다.

## 3. 관련연구

### 3.1 협력적 필터링

협력적 필터링은 사용자 집단의 정보에 기반 하여 사용자의 선호도를 예측하는데 널리 사용되는 정보 필터링 기술이다. 예를 들어, 사용자에게 의해 평가된 상품에 대한 점수를 알고 있을 때, 이 기법을 통하여 그 사용자가 평가하지 않은 어떤 상품이 어떤 선호도 점수를 가질 수 있는지를 예측할 수 있다. 사용자는 그와 비슷한 선호도를 가지는 사람들이 좋아하는 것을 좋아할 것이라고 가정하는 것이 이 기법의 생각이다. 이 기법의 적용은 Amazon.com 등의 상거래 사이트에서 적용되고 있는 사례이다. 대

부분의 협력적 필터링 시스템은 일정 범위의 수치 점수(평가점수)로 표현된 사용자의 의견을 수집한다. 수집된 점수들은 행렬 형태로 저장되고, 협력적 필터링 시스템은 이 점수 행렬을 사용하여 예측한다. 대개의 협력적 필터링 시스템이 수행하는 작업은 사용자  $u_i$ 의 상품(항목)  $i$ 에 대한 점수를 예측하는 것이다. 시스템은 사용자  $u_i$ 의 점수들을 기존의 상품(항목)  $i$ 에 선호도(평가점수)가 있는 다른 사용자들의 선호도(평가점수)와 비교한다. 그런 다음 다른 사용자들의 점수에 대한 가중치를 고려한 평균을 사용하여 예측이 이루어진다. 가중치는 사용자  $u_i$ 와 다른 사용자들 사이의 유사성을 수치로 표현한 것이며, 가중치가 고려된, 여러 사용자들의 상품(항목)  $i$ 에 대한 점수 평균이 사용자  $u_i$ 가 상품(항목)  $i$ 에 대한 점수 평균이 사용자  $u_i$ 가 상품(항목)  $i$ 에 대해 평가할 것이고 예측하는 선호도(평가점수)이다.

### 3.2 내용기반 필터링

상품(항목)들은 각각의 여러 속성을 가진다. 한 사용자에게 있어, 이전에 구매된 상품과 비슷한 속성을 가지는 상품은 그의 취향에 맞을 가능성이 높다. 상품들의 속성에 있어 비슷한 정도를 사용하여 상품 유사도를 계산하는 것이 내용 기반 필터링이다. (콕미라(2002))

### 3.3. 선행연구

이용준(2003)은 협업여과방법(협력적여과방법,협동적필터링,collaborative filtering)의 문제점으로 지적되고있는 희소성(sparsity)으로 인한 유사도의 부정확한 문제를 개선하기 위하여 인구통계정보를 이용한 기법을 제안하였으며, 박지선(2001)은 협동적 필터링 기술에서 적용하고있는 피어슨 상관계수를 이용하는 방법에서 비슷한 선호도 패턴을 가지는 고객들을 적절히 군집화하여 이 군집에 속하는 고객들의 평가를 기반으로 하여 협동적 필터링 기술을 수행하는 방법을 제안하였으며, 김병만(2004)은 추천시스템의 예측방법으로 내용기반 필터링(content based filtering)과 협력필터링(collaborative filtering)의 결합기법을 사용하였다.김택현(2004)는 ordered clustering 방법을 이용하여 each movie dataset에 대해 예측정확도를 실험하였다. 또한 박지선(2002)은 예측능력을 향상시키기 위해 유사한 선호도를 가지는 고객들의 평가에 근거하여 상품들간의 유사도를 구하여 특정상품에 대한 고객의 선호도를 예측하는 추천기법을 제안하였으며 김영철(2002 a)은 grouplens의 movielens dataset에 대해 영화 장르를 이용하여 예측의 오차를 줄이는 방법에 대해 연구하였으며 콕미라(2001)는 협력적 필터링 방법에 내용기반 필터링 방법을 적용하고 고객의 상품에 대한 선호도를 계산하는 방법을 제안하였다. 김종수(2001)는 신경망을 이용한 협력적 추천시스템의 정확도를 향상시키는 방법을 제안하였다.

#### (1) 사용자 기반 협동적 필터링

사용자 기반 협동적 필터링은 특정 고객의 상품에 대한 선호도를 예측하기 위하여 대부분의 경우 식 (2)에 나타나 있는 피어슨 상관 계수를 이용하여 선호도(유사도) 계

산하고 식 (1)에 의해 예측 선호도 값을 계산한다.

$$U_x = \bar{U} + \frac{\sum_{J \in \text{Raters}} (J_x - \bar{J}) r_{uj}}{\sum_{J \in \text{Raters}} |r_{uj}|}, \quad (1)$$

여기서

$$r_{uj} = \frac{\sum (U - \bar{U})(J - \bar{J})}{\sqrt{\sum (U - \bar{U})^2} \sqrt{\sum (J - \bar{J})^2}}, \quad -1 \leq r_{uj} \leq 1 \quad (2)$$

$U_x$ 는 상품  $x$ 에 대한 고객  $U$ 의 예측된 선호도이고,  $r_{uj}$ 는 고객  $U$ 와  $J$ 의 상관관계를 나타내며 두 고객 모두 선호도를 표시한 상품에 대해서만 계산된다. 여기서  $J_x$ 는 상품  $x$ 에 대한 고객  $J$ 의 선호도를 나타내며  $\bar{J}$ 는 고객  $J$ 의 평균 선호도를 의미한다.  $r_{uj}$ 가 1에 가까울수록 두 고객의 선호도 경향이 매우 유사함을 나타내고 -1에 가까울수록 반대의 선호 경향을 나타낸다. Raters는 테스트 상품에 대해 선호도를 표시한 고객들을 의미한다.(박지선(2002))

#### (2) 예측의 정확성에 대한 평가 기준

본 논문에서는 예측 값의 정확성을 평가하기 위해 MAE(Mean Absolute Error)를 사용하였으며 식 (3)에 나타낸 것과 같이 구할 수 있다.

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

## 4.. 연구 가설 및 평가 자료의 분석

### 4.1 연구가설

본 논문에서의 연구 가설은 다음과 같이 설정 하였다.

- (1) 연구가설1 : 성별에 따른 예측 시 예측오차의 추정치는 감소할 것이다.  
movielens dataset에서 제공되는 인구통계 변수 중 성별자료를 사용하여 예측을 하면 예측 오차를 감소시킬 것 이라는 가정에서 연구가설1을 설정하였다.
- (2) 연구가설2 : 연령에 따른 추정 예측치는 감소할 것이다.  
movielens dataset에서 제공되는 인구통계 변수 중 연령자료를 사용하여 예측을 하면 예측 오차를 감소시킬 것 이라는 가정에서 연구가설2를 설정하였다.
- (3) 연구가설3 : 응답 쌍에 따른 추정 예측치는 감소할 것이다.  
movielens dataset에서 제공되는 자료의 희소성(sparsity)의 원인으로 예측 오차가 증가 할 수 있으므로 응답편수에 따른 추정 예측치와 예측오차는 관련성이 있다는 가정 하에 연구가설3을 설정하였다.
- (4) 연구가설4 : 상관계수의 크기에 따라 추정 예측치는 영향이 있을 것이다.

예측의 오차는 상관계수의 크기에 따라 차이가 있을 것이라는 가정 하에 연구가설4를 설정하였다. 한편 직업에 따른 예측은 사전조사 결과 응답 쌍의 희소성으로 예측하기에 가능하지 않아 가설로 설정하지 않았다.

#### 4.2 평가자료의 분석

본 논문에서 사용된 dataset은 GroupLens의 movielens dataset을 이용하여 실험을 하였다. movielens dataset은 943명의 평가자들은 1682편의 영화에 대해 최소 20편을 평가하였으며 최대 737편의 영화에 평가를 하였다. 평가점수는 1-5점으로 평가하였다. 1682편의 영화에 943명이 평가한 평가의 수는 100,000개이다. 본 논문에서는 GroupLens에서 제공되는 movielens dataset중 u1.base, u2.base, u3.base, u4.base, u5.base를 training dataset으로 u1.test, u2.test, u3.test, u4.test, u5.test를 test dataset으로 사용하였다.u1.base,...u5.base와 u1.test,...,u5.test는 전체 응답 자료를 서로 중복이 되지 않도록 분할 되어있으며 training dataset과 test dataset이 각각 80%, 20%로 나뉘어져 있다.

##### (1) 응답 평가자의 성별분포

movielens dataset의 응답 평가자의 성별분포는 다음 <표 1>과 같다.

<표 1> 성별의 분포표

성별	빈도수	비율
남자	670	71.05
여자	273	28.95
계	943	100.0

평가자의 성별분포는 남자가 71.05%, 여자가 28.95%로 응답하였다.평가자의 성별에 따른 예측치는 전체적인 예측오차에 비해 낮아질 것이라는 기대감이 있으며 성별을 고려한 예측치의 추정이 필요할 것으로 기대된다.

##### (2) 응답 평가자의 연령분포

movielens dataset의 응답 평가자의 연령분포는 다음 <표 2>와 같다.

<표 2> 응답 평가자의 연령분포표

연령	빈도수	비율
10대이하	77	8.17
20-29세	332	35.21
30-39세	241	25.56
40-49세	168	17.82
50세 이상	125	13.26
계	943	100.0

평가자의 연령분포 10대이하가 8.17%, 20-29세가 35.21%, 30-39세가 25.56%, 40-49세가 17.82%, 50세이상인 13.26%로 나타났다. 성별에 따른 예측이 필요한 것처럼 연령에 따른 요인이 예측의 오차를 크게 할 수 있을 것이므로 연령을 고려한 예측치의 추정이 필요할 것으로 기대된다.

(3) 응답 평가자의 성별에 따른 연령분포표

movielens dataset의 응답 평가자의 성별에 따른 연령분포는 다음 <표 3>과 같다.

<표3> 성별에 따른 연령의 교차표

구 분			연령별					
			10대이하	20대	30대	40대	50대이상	전체
성별	남	빈도	45	247	175	112	91	670
		Row 비율	6.7%	36.9%	26.1%	16.7%	13.6%	100.0%
	녀	빈도	32	85	66	56	34	273
		Row 비율	11.7%	31.1%	24.2%	20.5%	12.5%	100.0%
	전체	빈도	77	232	241	168	125	943
		Row 비율	8.2%	35.2%	25.6%	17.8%	13.3%	100.0%

(4) 응답 평가자의 응답편수의 빈도 분포

movielens dataset의 응답 평가자의 응답편수의 빈도 분포는 다음 <표 4>와 같다.

<표 4> 응답편수의 빈도 분포표

응답편수	빈도	퍼센트
20-25	137	14.5
26-30	76	8.1
31-40	92	9.7
41-60	144	15.3
61-80	87	9.2
81-100	46	4.2
101-150	435	13.9
151-200	78	8.3
200이상	148	15.7

<표 4>의 응답분포표를 분석하면 movielens 자료는 943명의 응답자가 1682편의 영화 중최소 20편의 영화에 응답하였다. 응답자는 평균 106.0개의 영화에 대해 평가를 하였으며 표준편차는 100.93으로 나타났다.

(5) 응답쌍(pair of response)의 빈도분포

응답쌍(pair of response)의 빈도분포표는 다음 <표 5>와 같다.

<표 5> 응답쌍(pair of response)의 빈도 분포표

응답쌍의 수	빈도	퍼센트
0	15043	3.4
1-5	125478	28.2
6-10	93886	21.2
11-15	57135	12.8
16-20	34846	7.9
21-25	23389	5.3
26-30	16585	3.7
31-50	37329	8.4
51이상	39235	9.1

movielens dataset의 응답 평가자의 응답편수의 응답쌍의 빈도분포표에서 알 수 있듯이 응답쌍이 없는 경우가 3.4%이며 응답쌍이 5개 이하의 경우 31.6%를 차지하고 있어 이상점(outlier)이 있을 경우 상관계수는 이상점의 영향을 받을 수 있다.

(6) 응답 점수(rating of response)의 빈도분포

응답 평가자의 응답 점수(rating of response)의 빈도분포는 다음 <표 6>과 같다.

<표 6> 응답 점수의 분포

평가점수	빈도수	비율
1	6110	6.1
2	11370	11.37
3	27145	27.15
4	34174	34.17
5	21201	21.20
계	100,000	100.0

<표 6>의 응답자 점수의 분포에서 평가점수의 분포를 보면 평가점수는 보통보다 높은 평가를 한 것으로 나타났다.

(7) MAE 결과

수식 (1)과 (2)에 의해 u1.base,...,u5.base를 training dataset으로 u1.test,...,u5.test를 test dataset으로 하여 예측치를 구한 결과 MAE(Mean Absolute Error)는 다음 <표 7>과 같이 나타났다.

&lt;표 7&gt; MAE(Mean Absolute Error)

구분	u1.base	u2.base	u3.base	u4.base	u5.base	비고
무조건MAE	0.7655	0.6263	0.6187	0.6150	0.6173	

<표 7>의 MAE(Mean Absolute Error)는 0.6150-0.7655로 나타났다.

### 4.3 가설의 검증

(1) 연구가설1 : 성별에 따른 예측 시 예측오차의 추정치는 감소할 것이다. movielens dataset에서 제공되는 인구통계 변수 중 성별자료를 사용하여 예측을 하면 예측 오차를 감소시킬 것 이라는 가정 하에서 u1.base,..u5.base를 training dataset으로 u1.test,..u5.test를 test dataset으로 예측을 하였다. test dataset의 평가자와 동일한 성별을 training dataset에서 고른 후에 예측을 하였다. 예측에 따른 MAE를 구한 결과는 다음 <표 8>과 같이 나타났다.

&lt;표 8&gt; 성별에 따른 MAE(Mean Absolute Error)

구분	u1.base	u2.base	u3.base	u4.base	u5.base	비고
성별MAE	0.7742	0.6387	0.6288	0.6271	0.6268	
무조건MAE	0.7655	0.6263	0.6187	0.6150	0.6173	

<표 8>에서 알 수 있듯이 성별에 따른 예측의 결과 MAE의 감소는 나타나지 않았음을 알 수 있다.

(2) 연구가설2 : 연령에 따른 추정 예측치는 감소할 것이다. movielens dataset에서 제공되는 인구통계 변수 중 연령자료를 사용하여 예측을 하면 예측 오차를 감소 시킬 것 이른 가정 하에 5개의 training dataset에 대해 5개의 test dataset으로 예측을 하였다. test dataset의 평가자와 동일한 연령을 training dataset에서 고른 후에 예측을 하였다. 예측에 따른 MAE를 구한 결과는 다음 <표 9>와 같이 나타났다.

&lt;표 9&gt; 연령에 따른 MAE(Mean Absolute Error)

구분	u1.base	u2.base	u3.base	u4.base	u5.base
연령MAE	0.7912	0.6538	0.6500	0.6457	0.6454
무조건MAE	0.7655	0.6263	0.6187	0.6150	0.6173

<표 9>에서 알 수 있듯이 연령에 따른 예측의 결과 MAE의 감소는 나타나지 않았음을 알 수 있다.



(3) 연구가설3 : 응답쌍에 따른 추정 예측치는 감소할 것이다.  
movielens dataset에서 제공되는 자료의 희소성(sparsity)의 원인으로 예측 오차가 증가 할 수 있으므로 응답편수에 따른 추정 예측치와 예측오차는 관련성이 있다는 가정 하에 5개의 training dataset에 대해 5개의 test dataset으로 예측을 하였다. training dataset의 응답쌍에 따라 예측오차는 어떻게 나타나는지를 분석하기 위해 응답 쌍을 몇가지로 나누어 예측을 하였으며 그 결과는 다음 <표 10>과 같다.

<표 10> 응답 쌍에 따른 MAE

응답쌍 dataset	3이상	5이상	7이상	10이상	20이상	30이상	50이상	70이상	무조건 MAE
u1.base	0.7639	0.7612	0.7603	0.7606	0.7527	0.7549	0.7566	0.7758	0.7655
u2.base	0.6269	0.6294	0.6331	0.6395	0.6621	0.6735	0.6909	0.7096	0.6263
u3.base	0.6192	0.6219	0.6258	0.6312	0.6554	0.6717	0.6858	0.7014	0.6187
u4.base	0.6156	0.6186	0.6223	0.6286	0.6549	0.6672	0.6832	0.7004	0.6150
u5.base	0.6181	0.6208	0.6240	0.6303	0.6542	0.6695	0.6885	0.7031	0.6173

<표 10>에서 알 수 있듯이 응답 쌍의 크기에 따른 예측의 결과 MAE의 감소는 나타나지 않았음을 알 수 있다.

(4) 연구가설4 : 상관계수의 크기에 따라 추정 예측치는 영향이 있을 것이다.  
예측의 오차는 상관계수의 크기에 따라 차이가 있을 것이라는 가정 하에 5개의 training dataset에 대해 5개의 test dataset으로 예측을 하였다. test dataset의 사용자와 영화에 대해 training dataset의 상관계수의 크기를 분류 한 후 예측을 하였다. 예측에 따른 MAE를 구한 결과는 다음 <표 11>과 같이 나타났다.

<표 11> 상관계수의 크기에 따른 MAE(Mean Absolute Error)

상관계수 dataset	0.2  이상	0.4  이상	0.6  이상	0.8  이상	무조건MAE
u1.base	0.7662	0.7857	0.8529	0.9398	0.7655
u2.base	0.6056	0.5723	0.5800	0.6058	0.6263
u3.base	0.5980	0.5682	0.5746	0.5946	0.6187
u4.base	0.5951	0.5628	0.5728	0.5929	0.6150
u5.base	0.5962	0.5654	0.5764	0.5954	0.6173

<표 11>의 상관계수의 크기에 따른 MAE를 구한 결과 u1.base를 제외한 나머지 4개의 dataset의 MAE는 상관계수의 크기에 영향을 받는 것으로 나타났다. 상관계수  $r=|0.4|$  일 때가 가장 낮은 MAE를 가지는 것으로 나타났다. 상관계수의 크기가  $|0.6|$ ,  $|0.8|$ 로 커질 때는  $|0.4|$ 보다 MAE가 증가하는 것은 자료의 희소성이 증가하므로 예측에 영향을 미치는 것으로 분석되었다. 상관계수가  $|0.2|$ 이하일 때는 예측의 MAE를 증가시키므로 예측에 안 좋은 영향을 미치는 것으로 분석할 수 있다.

## 5. 결 론

이 논문에서 설정한 연구가설중 성별에 따른 추정 예측치는 감소하지 않았으며, 연령에 따른 예측에서도 추정 예측치는 감소하지 않았다. 또한 응답편수에 따른 추정 예측치도 감소하지 않았으나 상관계수의 크기에 따른 추정 예측치는 감소하였다. 이 논문에서 사용된 Grouplens의 movielens dataset은 자료의 희소성을 가지는 자료이다. 사용자 기반 추천시스템의 예측에서 상관계수의 크기를 고려한 예측시스템이 필요한 것으로 분석되었다. 본 논문에서는 유의적인 상관계수가 예측 오차에 어떤 영향을 주는지는 분석하지 못하였다. 앞으로의 연구 방향은 MAE 오차를 감소하는 방법과 유의적인 상관계수가 오차에 어떤 영향을 주는지의 연구가 필요하며 MAE의 감소와 독립적으로 평가순위의 일치성에 대한 연구가 필요할 것으로 기대된다.

## 참고문헌

1. 광미라, 조동섭(2001), 개인화된 추천 시스템의 선호도 계산을 위한 정보 필터링, 한국정보과학회 2001년 춘계학술대회, 28권, 1호.
2. 김병만, 이경, 김희관, 임은기, 김주연(2004), 추천시스템을 위한 내용기반 필터링과 협력필터링의 새로운 결합 기법, 한국정보과학회 논문지 B - 소프트웨어 및 응용, 31권, 3호.
3. 김영설, 김병천, 윤병주(2002), 개선된 추천시스템을 이용한 전자상거래시스템 설계 및 구현, 정보처리학회논문지 D, 9-D권, 2호.
4. 김종수, 도영아, 류정우, 김명원(2001), 신경망을 이용한 협력적 추천 시스템의 성능 향상, 한국뇌학회지, 1권, 2호.
5. 김택현, 양성봉(2004), 협동적 필터링 기반 추천 시스템을 위한 향상된 이웃 선정 방법, 제 21회 한국정보처리학회 춘계학술발표대회, 11권, 1호.
6. 박지선, 김택현, 유영석, 양성봉(2001), 추천 시스템을 위한 고객 클러스터링 방법을 적용한 예측 알고리즘, 한국정보과학회 2001년 춘계학술대회, 28권, 1호.
7. 박지선, 김택현, 유영석, 양성봉(2002), 추천 시스템을 위한 2-way 협동적 필터링 방법을 이용한 예측 알고리즘, 한국정보과학회 논문지 B - 소프트웨어 및 응용, 29권, 9호.
8. 이용준, 이세훈, 왕창중(2003), 인구 통계 정보를 이용한 협업 여과 추천의 유사도 개선 기법, 한국정보과학회 논문지 C - 컴퓨팅의 실제, 9권, 5호.

[ 2005년 11월 접수, 2006년 2월 채택 ]