

협력적 필터링 알고리즘의 예측 선호도 순위 일치와 Top-N 추천에 관한 연구

(A Study on the Relation of Top-N Recommendation
and the Rank Fitting of Prediction Value through a
Improved Collaborative Filtering Algorithm)

이석준*, 이희춘**

(Seok-Jun Lee, Hee-Choon Lee)

요 약 본 연구는 추천시스템에서 협력적 필터링 알고리즘인 이웃기반의 협력적 필터링 알고리즘과 대응평균 알고리즘의 선호도 예측 결과를 이용하여 예측결과와 순위 일치성과 실제 고객에 상품 추천인 Top-N 추천의 정확도에 대하여 연구하였다. 연구결과 대응평균 알고리즘에 의한 선호도 예측 정확도의 순위 일치성과 예측치를 이용한 Top-N 추천의 정확도가 기존의 이웃기반의 협력적 필터링 알고리즘의 결과보다 우수함을 알 수 있었다. 이는 협력적 필터링 추천시스템에서 대응평균 알고리즘을 이용한 선호도 예측 결과를 이용하여 고객에게 상품추천을 하는 것이 이웃기반의 협력적 필터링 알고리즘을 이용하는 것보다 더 효과적이며 추천시스템에 대한 고객의 만족을 향상 시킬 수 있을 것으로 기대된다.

핵심주제어 : 협력적 필터링, 추천시스템, 대응평균 알고리즘, Top-N 추천

Abstract This study devotes to compare the accuracy of Top-N recommendations of items transacted on the web site for customers with the accuracy of rank conformity of the real ratings with estimated ratings for customers' preference about items generated from two types of collaborative filtering algorithms. One is Neighborhood Based Collaborative Filtering Algorithm(NBCFA) and the other is Correspondence Mean Algorithm(CMA). The result of this study shows the accuracy of Top-N recommendations and the rank conformity of real ratings with estimated ratings generated by CMA are better than that of NBCFA. It would be expected that the customer's satisfaction in Recommender System is more improved by using the prediction result from CMA than NBCFA, and then Using CMA in collaborative filtering recommender system is more efficient than using NBCFA.

Key Words : Collaborative Filtering, Correspondence Mean Algorithm, Recommender System, Top-N Recommendation

1. 서 론

인터넷의 발전으로 등장한 전자상거래는 전통적 상거래의 시간과 공간의 제약을 해결하여 상거래 비용의 절감과 같은 이득을 실현하고 있다. 그러나 전자상거래 규모의 확대와 다양한 유형의 전자상

* 상지대학교 경영학과 겸임교수

** (교신저자) 상지대학교 컴퓨터데이터정보학과 교수

거래가 활성화됨에 따라 전자상거래에서 거래되는 상품에 대한 수많은 정보들이 양산되고 있으며 이러한 정보의 생성은 곧 고객에게 자신이 선호하는 상품을 검색하고 타 상품과 비교하고 구매하는 일련의 구매행위에 시간과 노력을 요구하게 되었다. 이러한 정보 과잉의 문제를 해결할 수 있는 대안으로 추천시스템이 제시되었으며 현재 많은 전자상거래 사이트들이 추천시스템을 활용하여 이러한 정보 과잉문제를 해결하고 구매력을 증대시키기 위해 노력하고 있다. 특히 Amazon.com과 같이 수많은 상품들을 판매하는 전자상거래 기업의 경우 추천시스템은 사이트에 대한 고객의 만족도를 높여 충성도와 구매력을 증대시킬 수 있는 중요한 시스템으로 운영되고 있다[1]. 추천시스템은 웹상에서 서비스, 정보, 상품등과 같은 아이템에 대한 고객의 선호 정보를 이용하여 자동적으로 새로운 아이템에 대한 선호도를 예측할 수 있는 시스템이다. 자동적으로 고객의 선호 정보를 생성할 수 있는 기술인 추천시스템은 고객의 정보검색 비용의 절감과 자신이 알지 못했던 새로운 정보를 제공받을 수 있는 다양한 이득이 있으며 시스템에서 얻어지는 이득을 통한 고객 서비스 만족도를 높일 수 있기 때문에 정보 과잉을 해결하기 위한 다양한 영역에서 연구가 이루어지고 있으며 적용되고 있다.

2. 연구목적

본 연구는 협력적 필터링에 의한 개별 사용자들의 선호도 예측 결과와 예측 결과를 이용한 상품 추천 정확도와 예측 결과의 순위관계에 대한 연구이다. 일반적으로 추천시스템의 성능은 선호도 예측 알고리즘의 예측 정확도를 이용하여 평가하고 있다[2,3,4,5]. 특히 본 연구에서 사용하는 MovieLens dataset과 같이 명시적 선호도 평가치의 경우 일반적으로 선호도 예측 알고리즘으로 생성된 예측치와 실제 dataset의 선호도 평가치와의 오차를 이용하여 알고리즘의 정확성을 평가한다. 또한 추천시스템에서 사용자에게 추천된 상품에 대한 추천은 예측 알고리즘으로 생성된 예측치를 바탕으로 사용자에게 적합한 N개수의 상품 목록을 작성한

다. 사용자들에게 일정 N개의 상품 추천을 Top-N 추천이라 한다. 이때 Top-N 추천의 상품목록은 선호도 예측 결과가 우수하게 평가된 상품으로 구성된다[6]. 기존의 연구에서 Top-N 추천 정확도에 대한 연구는 알고리즘으로 생성된 선호도 예측치의 예측 오차가 작으면 Top-N 추천의 정확도가 향상되는 것으로 분석하였다[7]. 본 연구에서는 협력적 필터링 기법의 두 가지 선호도 예측 알고리즘에 의해 생성된 선호도 예측의 결과를 이용한 Top-N 추천의 정확도를 오차의 개념이 아닌 선호도 예측치의 순위 개념을 이용하여 통계적으로 분석하였다. 알고리즘의 성능을 정확하게 테스트하기 위해 기존에 알고리즘의 제시되던 training set과 test set의 구분을 통한 예측 정확도의 평가 방법이 아닌 전체 dataset의 모든 평가치들에 대한 선호도 예측을 실시하여 선호도 예측치의 순위와 Top-N 추천정확도와와의 관계를 분석하였다. 결국 본 연구의 목적은 Top-N 추천의 정확도를 높이기 위해서는 알고리즘을 통해 생성된 선호도 예측치와 실제 평가치의 선호도 예측 오차보다 순위일치의 중요성을 밝히고 선호도 예측 오차가 크더라도 선호도 예측치의 순위와 실제 평가치의 순위의 일치가 잘 이루어지는 선호도 예측 알고리즘이 실제 추천시스템에서 더 효율적임을 연구하였다.

3. 선행연구

3.1 협력적 필터링

협력적 필터링 기법의 추천시스템은 목표 상품을 제안하기 위해 시스템 내의 고객들이 이전에 상품에 대해 선호도를 평가한 자료를 기반으로 목표 고객 자신의 선호도와 유사한 성향의 고객들의 정보를 이용하여 목표 상품에 대한 선호 정도를 예측한다. 그래서 협력적 필터링 기법을 “사회적” 필터링이라고도 부르며 실제 많은 전자상거래 사이트에 적용되고 있는 기법이다[3]. 상품의 선호도 예측은 다른 고객들이 목표 상품에 대하여 선호 정도를 평가한 수치적 평가치를 바탕으로 계산되는데 이 때 예측을 위한 다른 고객이 추천 대상 고객의 이웃이 되며 이들 이웃과의 선호정도를 계

산하여 예측한다[8]. 협력적 필터링에서 고객과 상품 간의 상호관계 데이터는 <그림 1>과 같이 행렬 형태로 표현하고 분석한다.

고객 \ 상품	User1	User2	User3	User4
Item1	$R_{1,1}$	$R_{2,1}$	$R_{3,1}$	
Item2	$R_{1,2}$	$R_{2,2}$		$R_{4,2}$
Item3		$R_{2,3}$	$R_{3,3}$	$R_{4,3}$
Item4	$R_{1,4}$		$R_{3,4}$?
Item5	$R_{1,5}$	$R_{2,5}$	$R_{3,5}$	$R_{4,5}$

User1	User3
$R_{1,1}$	$R_{3,1}$
$R_{1,2}$	
	$R_{3,3}$
$R_{1,4}$	$R_{3,4}$
$R_{1,5}$	$R_{3,5}$

(그림 1) 협력적 필터링에서 고객과 상품과의 행렬

(그림 1)에서 고객4(user4)의 상품4(item4)에 대한 선호도를 예측하기 위해서는 선호도 예측에 사용될 이웃 사용자들의 선호도 평가치가 필요하다. (그림 1)에서는 고객1과 고객3이 고객4의 선호도 예측을 위해 필요한 이웃 사용자가 된다. 협력적 필터링은 선호도 예측치를 생성하기 위한 사용자 자신의 선호 정보와 예측 대상이 되는 상품에 이미 평가한 이웃 사용자들의 선호 정보를 동시에 고려하게 된다. 이러한 개념의 협력적 필터링 접근법은 상업적으로 가장 널리 이용되고 성공적인 추천 접근법으로 알려져 있으며 추천 알고리즘 연구의 근간을 이루고 있다[2,9,10].

협력적 필터링 알고리즘은 메모리 기반(memory based)의 알고리즘과 모델 기반(memory based)의 알고리즘으로 구분한다[11]. 메모리 기반의 알고리즘은 사용자들이 상품에 대해 평가한 전체 선호도 평가치를 기반으로 선호도 평가치에 대한 예측이 이루어지며 특정 상품에 대한 추천 대상 사용자의 선호도를 예측하기 위하여 데이터베이스의 모든 평가치를 이용하게 된다. 반면 모델 기반의 알고리즘은 사용자들의 평가치 특성에 따라 소규모 클래스로 구분하여 클래스별로 목표 상품에 대한 선호도를 예측하는 방법으로 통계적 기법들이 이용된다. 메모리 기반의 알고리즘은 상품과 사용자의 수가 증가하면 알고리즘을 통한 계산량이 증가한다는 단점을 가지고 있지만 모델 기반의 알고리즘에 의해 생성된 예측치와 비교하여 다양한 선호도 예측치를 제공하며 이는 예측치의 선호 순위를 구분하기에 적합하다. 일반적으로 선호도 평가치 기반

의 추천시스템에는 메모리 기반의 알고리즘이 적용된다. Herlocker 등(2004)의 연구에서는 이웃 기반의 협력적 필터링 알고리즘(Neighborhood Based Collaborative Filtering Algorithm)을 이용한 추천시스템의 성능을 향상시키기 위하여 선호도 예측 과정에서 이웃의 수에 따른 예측 정확도의 개선과 공통으로 평가된 상품의 수를 고려한 유의성 가중치에 따라 예측의 정확도가 향상됨을 보이고 있다. 또한 예측 오차의 경우 오차의 절대 평균이 0.73 이하로 개선시키기 어렵다고 밝히고 있으며 추천시스템 성능 분석에 대한 다양한 가능성에 대하여 제시하고 있으나 구체적인 분석결과는 제시되지 않았다[12]. 또한 이희춘의 연구에서도 선호도 예측 알고리즘으로 생성된 선호도 예측치와 실제 선호도 평가치의 예측 오차가 작으면 Top-N 추천의 정확도도 높아짐을 연구하였지만 실제 Top-N 추천의 경우 예측 오차보다 선호도 예측치의 순위와 실제 평가치의 순위의 적합성에 따라 정확도가 향상된다[7]. 기타 다양한 범주의 알고리즘이 협력적 필터링 기법에 제안되었으며 선호 경향이 유사한 이웃의 구성과, 분류를 위한 알고리즘, 연관규칙 마이닝, 베이시안 네트워크, 그리고 군집화 모형과 같은 많은 데이터 분석 알고리즘들이 협력적 필터링 문제에 적용되었다. 김경재, 김병국 (2005)은 유전자 알고리즘을 이용하여 고객 정보에서 성향을 추출할 수 있는 추천시스템의 추천엔진 개발에 대해 연구하였다[13]. 김제경 등(2003)은 거래기록에서 상품과 고객 간의 연관규칙에 대한 패턴을 끌어내기 위하여 연관규칙 마이닝 기법이 추천시스템에 적용되었다[14]. 진승훈 등(2002)은 신경망을 이용한 유즈넷 뉴스의 필터링에 대하여 연구하였다[15]. 본 연구에서는 협력적 필터링 알고리즘 중 선호도 평가치 기반의 시스템에 일반적으로 적용되는 이웃 기반의 협력적 필터링 알고리즘(NBCFA)과 이를 개선하여 예측 성능이 향상된 대응 평균 알고리즘(Correspondence Mean Algorithm)의 성능을 선호도 예측 오차의 개념이 아닌 선호도 예측치와 실제 평가치의 순위의 적합성을 비교하여 추천 성능을 분석하였다.

3.2 유사도 가중치

협력적 필터링에서 추천 대상 고객과 이웃 고객의 선호도 유사정도를 구하기 위하여 수치적 관계를 정의한다. 이때 두 고객 간의 선호도 유사 정도는 피어슨 상관계수, 벡터 유사도, 유클리디안 거리, 역사용자 빈도 등 다양한 형태의 가중치로 정의될 수 있으며 본 연구에서는 피어슨 상관계수를 이용하여 두 사용자의 선호도 유사정도를 정의하였다[2,10]. 다음 식(1)은 본 연구에서 유사도 가중치로 사용된 피어슨 상관계수이다.

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

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

3.3 예측 알고리즘

최초의 자동화된 협력적 필터링 알고리즘은 GroupLens에서 제시되었으며 추천 대상 고객과 이웃고객들의 선호도 평가치를 이용하여 선호도를 예측한다[2,16]. 다음 식(2)는 GroupLens에서 제안한 이웃 기반의 협력적 필터링 알고리즘(Neighborhood Based Collaborative Filtering Algorithm)이다.

$$\hat{U}_x = \bar{U} + \frac{\sum_{J \in \text{Raters}} (J_x - \bar{J})r_{uj}}{\sum_{J \in \text{Raters}} |r_{uj}|} \quad (2)$$

여기서 \hat{U}_x 는 목표 상품 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 목표 상품 x 에 대한 이웃고객 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 식(1)에서 정의된 피어슨 상관계수를 사용한다. Lee (2006)의 연구에서는 GroupLens에서 제안한 NBCFA알고리즘을 개선한 대응평균 알고리즘(Correspondence Mean Algorithm)을 제안하였으며 NBCFA에 의한 선호도 예측결과보다 예측 오차에 있어 우수한 성능을 나타내고 있음이 연구되었다[17,18]. CMA는 다음 식(3)과 같이 정의한다.

$$U_x = \bar{U}_{match} + \frac{\sum_{J \in \text{Raters}} (J_x - \bar{J}_{match})r_{uj}}{\sum_{J \in \text{Raters}} |r_{uj}|} \quad (3)$$

여기서, \bar{U}_{match} 는 \bar{U} 의 단점을 보완하기 위해 추천 대상 고객 u 와 각 이웃 고객 j 가 공통으로 평가한 상품에 대한 평가치의 평균들 즉, 고객 u 와 이웃고객 j 의 유사도 가중치를 계산할 때 이용되는 평가치들만을 이용한 평균들을 구하고 이 평균들을 다시 평균한 값이 된다. \bar{J}_{match} 역시 공통으로 평가된 아이템들에 대한 선호도 평가치의 평균으로 정의된다.

3.4 순위적합률

일반적으로 협력적 필터링 알고리즘의 선호도 예측 정확도를 평가하기 위해서 실제 선호도 평가치와 예측치 간의 오차 개념이 이용되지만 실제 추천 대상 고객에게는 선호도 예측치를 이용하여 추천 대상 고객이 가장 선호하리라 생각되는 N 개의 상품 목록을 추천하는 방법이 일반적으로 적용되고 있다[6]. 이때 추천 순위에 대한 정확도를 순위적합률이라 정의하여 평가한다. 순위적합률은 다음 식(4)와 같이 정의한다[19].

$$\frac{N(\text{Top}_N(R_u) \cap \text{Top}_N(\hat{R}_u))}{N(\text{Top}_N(\hat{R}_u))} \cdot 100 \quad (4)$$

- $N(Top_N(\widehat{R}_u))$: 목표 고객 u 의 선호도 예측치들인 \widehat{R}_u 에서 상위 N 개의 예측치들의 개수
- $N(Top_N(R_u) \cap Top_N(\widehat{R}_u))$: 예측된 선호도 상위 N 개의 목록과 동시에 발생하는 실제 선호도 상위 N 개의 상품 개수.

즉, 예측된 선호도 평가치 중 상위 N 개의 상품 목록과 비교하여 실제 선호도 평가치 중 상위 N 개에 해당하는 상품이 몇 개 발생했는지를 이용하여 추천 대상 고객에 대한 Top- N 추천의 정확도를 측정한다.

4. 실험

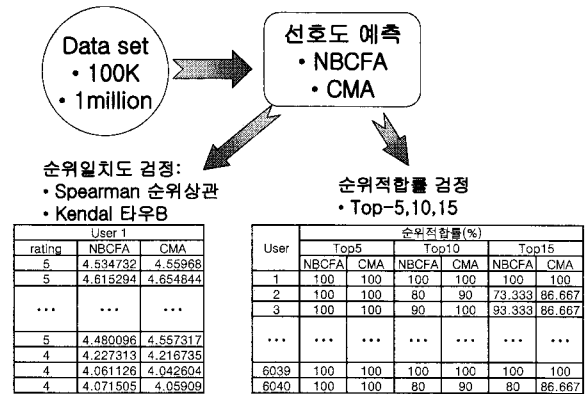
4.1 실험데이터

본 연구는 GroupLens에서 공개하는 100K MovieLens dataset과 1million MovieLens dataset을 이용하여 분석하였다. 100K MovieLens dataset은 총 943명의 사용자(user)와 1682편의 영화에 대해 그리고 1million dataset의 경우 총 6040명의 사용자와 3952편의 영화에 대해 사용자들이 자신의 영화에 대한 선호정도를 5점 척도로 표기한 선호도 평가치로 구성되어 있으며 각 사용자들은 최소 20편의 영화와에 평가하였으며 100K dataset에서는 943명 중 32명이 20편의 영화에 대해 평가하였고 최대 737편의 영화에 평가한 사용자도 있으며 1million dataset의 경우 최소 20편의 영화에 평가한 사용자는 86명이며 최대 2314편의 영화에 평가한 사용자도 있다. 또한 총 선호도 평가치의 개수는 100K dataset은 100,000개, 1million dataset은 1,000,209개로 구성되어 있으며 부가적으로 각 이용자에 대한 간단한 인구통계학적 정보와 영화의 장르에 대한 정보가 주어져 있다.

4.2 실험설계

본 연구의 실험은 100K dataset과 1million dataset의 모든 평가치에 대하여 두 선호도 예측 알고리즘으로 선호도 예측치를 생성하고 dataset의 개별 사용자인 943명과 6040명의 개인별 선호도 예

측치의 순위와 실제 개인별 선호도 평가치 순위의 일치도를 비교하였다. 또한 선호도 예측치를 이용한 Top- N 추천의 정확도인 순위적합률을 구하고 순위 일치도와와의 관계를 파악하였다. 선호도 예측치와 실제 선호도 평가치의 순위 일치도를 비교하기 위하여 Spearman 순위상관계수와 Kendall의 타우-b를 이용하여 순위 일치성을 통계적으로 비교하였다. 다음 (그림 2)는 실험 과정을 나타내고 있다. 순위일치도 검정에서는 개별 사용자의 실제 선호도 평가치와 NBCFA와 CMA에 의한 예측치의 관계를 나타내고 있다. 그리고 순위적합률 검정에서는 각 개인별 선호도 예측치로 생성된 Top-5, 10, 15의 목록의 정확도를 나타내는 순위적합률의 알고리즘 별 차이에 대한 검정을 나타내고 있다.



(그림 2) 실험 과정

5. 실험결과

5.1 선호도 예측치의 순위 일치도 평가

NBCFA와 CMA의 선호도 예측 결과의 순위 일치도를 비교하기 위하여 100K dataset의 각 개인별 선호도 예측치와 실제 선호도 평가치의 순위에 대하여 Spearman 순위상관계수와 Kendall의 타우b를 계산하였다. 각 개인별 Spearman 순위상관계수와 Kendall의 타우b에 대하여 대응평균검증을 실시하여 두 알고리즘 결과의 차이를 검정하였다. 다음 (표 1)은 각 개인별 Spearman 순위상관계수와 Kendall의 타우b의 결과를 이용한 NBCFA(이웃기

반 협력적 필터링)와 CMA(대응평균 알고리즘)에 대한 대응평균검정 결과이다.

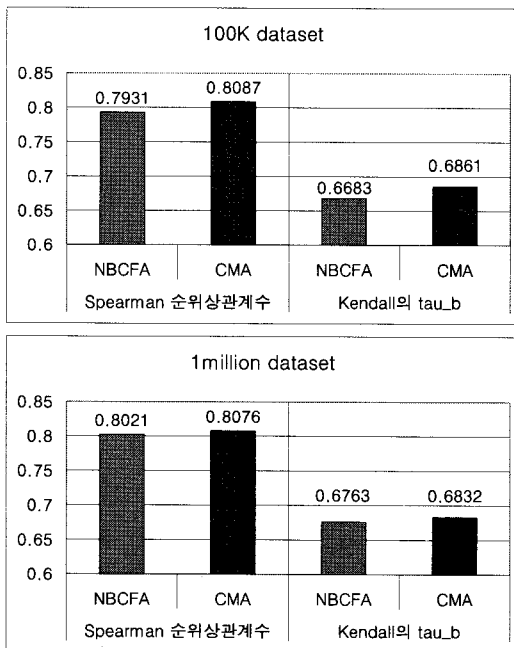
<표 1> 선호도 예측치와 실제 평가치의 순위 일치성에 대한 대응평균 검정결과

dataset	분석	평균		t값	유의 확률
		NBCFA	CMA		
100K	순위상관	0.7931	0.8087	-21.8954	0.000**
	타우B	0.6683	0.6861	-22.8197	0.000**
1million	순위상관	0.8021	0.8076	-29.5524	0.000**
	타우B	0.6763	0.6832	-32.2758	0.000**

*:p(0.05), **:p(0.01)

분석결과 100K dataset과 1million dataset에 대하여 NBCFA와 CMA에 의해 예측된 선호도 평가치와 실제 선호도 평가치의 순위에 대한 상관관계를 나타내는 Spearman 순위상관계수와 Kendall의 타우b에서 모두 CMA의 선호도 예측치 순위 일치가 NBCFA에 비하여 높게 나타남을 알 수 있다.

다음 (그림 3)은 100K dataset과 1million dataset의 Spearman 순위상관계수와 Kendall의 타우b의 결과의 각 알고리즘별 평균에 대한 그래프이다.



(그림 3) Spearman 순위상관계수와 Kendall의 타우b의 결과의 각 알고리즘별 평균

5.2 순위적합률의 평가

NBCFA와 CMA의 선호도 예측 결과를 이용하여 개인별 Top-N 추천의 정확도인 순위적합률을 Top-5, Top-10, Top-15에 대하여 구하고 알고리즘간 순위적합률의 차이를 대응평균검증을 통하여 분석하였다. 다음 (표 2)는 각 개인별 NBCFA와 CMA에 따른 Top-5, Top-10, Top-15 순위적합률의 대응평균검정 결과이다.

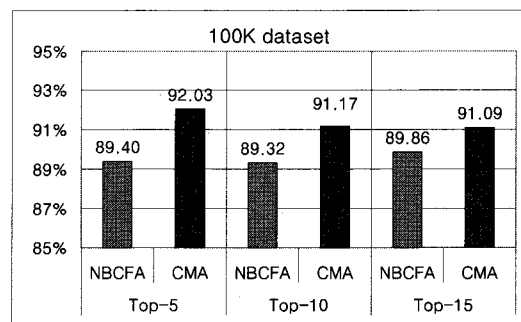
<표 2> 순위적합률에 대한 대응평균 검정 결과

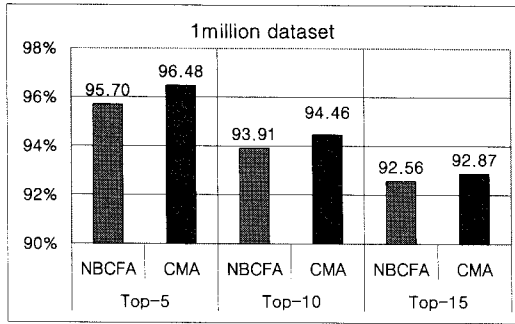
dataset	분석	평균		t값	유의 확률
		NBCFA	CMA		
100K	Top5	89.396	92.025	-7.51884	0.000**
	Top10	89.321	91.166	-8.65337	0.000**
	Top15	89.862	91.092	-7.56912	0.000**
1million	Top5	95.699	96.480	-9.0576	0.000**
	Top10	93.906	94.462	-8.5744	0.000**
	Top15	92.564	92.868	-5.4047	0.000**

*:p(0.05), **:p(0.01)

분석결과 100K dataset과 1million dataset에 의해 예측된 선호도 평가치와 실제 선호도 평가치에 대한 순위적합률은 모두 CMA의 순위적합률이 NBCFA에 비하여 높게 나타남을 알 수 있다. 다음

(그림 4)는 Top-5, Top-10, Top-15에 대한 순위적합률 평균에 대한 그래프이다.





(그림 4) Top5, Top10, Top15 순위적합률 결과의 각 알고리즘별 평균

6. 결론

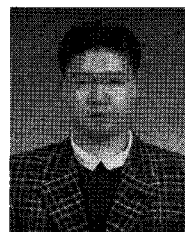
일반적으로 추천시스템의 성능에 대한 분석은 실제 선호도 평가치와 선호도 예측치의 오차의 개념을 이용하여 평가한다. 본 연구는 기존의 협력적 필터링 기법의 선호도 예측 알고리즘인 NBCFA와 이를 개선한 CMA에 의해 예측된 선호도 평가치의 예측 오차에 대한 연구와 달리 선호도 예측치의 순위와 실제 선호도 평가치의 순위 적합에 대하여 연구하였다. 기존 연구에서 NBCFA에 비하여 CMA의 예측 오차가 작게 나타남을 알 수 있었으나 본 연구는 예측 오차와 Top-N 추천 정확도의 관계에 대한 연구와 달리 NBCFA와 CMA에 의한 선호도 예측치의 순위가 어느 알고리즘이 우수한지에 대하여 연구하였다. 비록 선호도 예측을 통하여 예측 오차가 큰 결과를 갖더라도 예측 결과의 순위 적합이 높다는 것은 고객의 선호 상품의 순위를 결정하는데 중요한 근거가 될 수 있으며 이를 바탕으로 일반적인 추천방법인 Top-N 추천 방법과의 관계를 살펴보고 일정 수준의 N개 상품 이상에도 적용이 가능한지 살펴보았다. 실험결과에서 NBCFA에서 생성된 예측치의 순위보다 CMA에 의해 생성된 예측치의 순위가 실제 dataset의 평가치 순위와 비교하여 순위의 일치성이 더 높은 것으로 분석되었으며 또한 생성된 예측치를 이용한 Top-N 추천에서도 CMA에 의해 생성된 예측치를 이용한 순위적합이 더 잘 이루어지는 것으로 분석되었다. 또한 기존 연구의 결과에서 두 알고리즘의 선호도 예측의 예측 오차 개념의 정확도는 NBCFA의 결과와 비교하여 CMA의 결과가 더 우수한 것으로 평가되

었는데 본 연구의 결과를 통하여 선호도 예측치와 실제 평가치의 순위에 있어서도 CMA에 의해 생성된 선호도 예측 결과가 NBCFA의 예측 결과보다 더 우수한 성과를 나타내고 있음을 알 수 있었다. 이는 협력적 필터링 기법을 통한 추천에 있어 NBCFA의 선호도 예측 결과를 이용하는 것 보다 CMA의 예측 결과를 이용하는 것이 실제 추천에서 더 효과적임을 의미하며 추천시스템을 이용하는 고객에게 시스템에 대한 신뢰도와 만족도를 더 높일 수 있을 것으로 기대된다. 차후의 연구로는 추천 순위 일치도와 Top-N과는 어떤 관련성을 가지고 있는지를 분석하고 Top-N의 추천 정확도를 높일 수 있는 방법의 연구가 필요할 것으로 생각된다.

참고 문헌

- [1] G. Linden, B. Smith, J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", IEEE Internet Computing, Vol.7, Issue 1, pp.76-80, 2003.
- [2] J. S. Breese, D. Heckerman and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", In Proceedings of the Fourteenth Annual Conference on Uncertainty in Artificial Intelligence, pp.43-52, 1998.
- [3] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating "word of mouth", in Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co.: Denver, Colorado, United States, pp.210-217, 1995.
- [4] 이석준, 이희춘, "협업 필터링 추천에서 대응평균 알고리즘의 예측 성능에 관한 연구", Information Systems Review, 제9권, 제1호, pp.85-103, 2007
- [5] H. C. Lee, "Improved Algorithm for User Based Recommender System", Journal of the Korean Data & Information Society, Vol.17, No.3, pp.717-726, 2006.

- [6] M. Deshpande and G. Karypis, "Item-based top-N recommendation algorithms", *ACM Trans. Inf. Syst.*, Vol.22, No.1, pp.143-177, 2004.
- [7] 이희춘, 이석준, 정연수, "협력적필터링에서 개선알고리즘을 이용한 순위적합에 관한 연구", 2006년도 한국자료분석학회 추계학술대회, pp.103-111, 2006.
- [8] J. L. Herlocker, J. A. Konstan, and J. Riedl, "An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms", *Information Retrieval*, Vol.5, No.4, pp.287-310, 2002.
- [9] P. Resnick, 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.
- [10] P. Resnick and H. R. Varian, "Recommender systems", *Communications of the ACM*, Vol.40, No.3, pp.56-58, 1997.
- [11] Adomavicius, G., A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions", *IEEE Transactions on Knowledge and DATA Engineering*, Vol.17, No.6, pp.734-749. 2005.
- [12] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, "Evaluating Collaborative Filtering Recommender Systems", *ACM Transactions on Information Systems*, Vol. 22, No. 1, pp.5-53, 2004.
- [13] 김경재, 김병국, "데이터 마이닝을 이용한 인터넷 쇼핑몰 상품추천시스템", *한국지능정보시스템학회논문지*, 제11권, 제1호, pp.191-205, 2005.
- [14] 김재경, 안도현, 조윤희, "Development of a Personalized Recommendation Procedure Based on Data Mining Techniques for Internet Shopping Malls", *한국지능정보시스템학회논문지*, 제9권, 제3호, pp.177-191, 2003.
- [15] 진승훈, 김종완, 이승아, 김영순, 김병만, "코호넨 신경망을 사용한 유즈넷 뉴스 필터링 에이전트 구현", *한국산업정보학회논문지*, 제7권, 제5호, pp.21-28, 2002.
- [16] J. L. Herlocker, J. A. Konstan, A. Borchers, J. Riedl, "An algorithmic framework for performing collaborative filtering", in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, ACM Press: Berkeley, California, United States, pp.230-237, 1999.
- [17] H. C. Lee, "Improved Algorithm for User Based Recommender System", *Journal of the Korean Data & Information Society*, Vol.17, No.3, pp.717-726, 2006.
- [18] H. C. Lee, 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.
- [19] 이희춘, "추천시스템에서 Top-N추천을 위한 순위적합에 관한 연구", *Journal of the Korean Data Analysis Society*, Vol.8, No.6, pp.2597-2607, 2006.



이 석 준 (Seok-Jun Lee)

- 정회원
- 1996년 2월 : 상지대학교 산업공학과 (공학사)
- 1999년 8월 : 상지대학교 산업공학과 (공학석사)
- 2007년 2월 : 상지대학교 경영학과 (경영학박사)
- 2006년 3월 ~ 현재 : 상지대학교 경영학과 겸임교수
- 관심분야 : 전자상거래, 추천시스템, 데이터 마이닝, 무선 네트워크