

목표고객의 연령속성을 이용한 협력적 필터링 추천 시스템의 정확도 향상

이석환* · 박승헌*

*인하대학교 산업공학과

Accuracy improvement of a collaborative filtering recommender system using attribute of age

Seog-Hwan Lee* · Seung-Hun Park*

*Industrial Engineering of Inha University

Abstract

In this paper, the author devised new decision recommendation ordering method of items attributed by age to improve accuracy of recommender system. In conventional recommendation system, recommendation order is decided by high order of preference prediction. However, in this paper, recommendation accuracy is improved by decision recommendation order method that reflect age attribute of target customer and neighborhood in preference prediction. By applying decision recommendation order method to recommender system, recommendation accuracy is improved more than conventional ordering method of recommendation.

Keywords : Collaborative Filtering, Recommender System

1. 서 론

추천 시스템은 상품을 추천받을 고객인 목표고객의 원하는 상품을 정확하게 예측할 수 있어야 한다. 추천 시스템이 목표고객의 원하는 상품을 정확하게 예측하지 못하면 온라인 매장에 대한 목표고객의 만족도는 낮아지게 된다. 따라서 온라인 매장은 목표고객의 만족도를 높이기 위해 추천의 정확도가 높은 추천 시스템을 개발하여 사용해야만 한다.

본 연구의 목적은 협력적 필터링 추천 시스템의 추천 정확도를 향상시키는 것이다. 그 추천의 정확도를 향상시키기 위해 다음과 같은 방법을 제안한다.

추천의 정확도를 향상시키기 위해서는 목표고객에게 추천할 추천목록을 정확하게 생성해야한다. 기존 추천 시스템에서 추천목록은 상품의 예측 선호도가 높은 순

서로 추천순위를 결정한다. 여기서 예측 선호도는 목표고객의 실제 선호도와 차이가 있을 수 있기 때문에 예측 선호도만을 사용한 추천순위의 결정은 추천의 정확도를 저하시킬 수 있다. 따라서 예측 선호도 외에 목표고객의 연령속성을 이용하여 추천순위를 결정할 수 있는 방법을 제안한다.

본 연구에서는 앞에서 설명한 사항을 개선하여 추천의 정확도를 향상시킬 수 있는 방법을 다음과 같이 제안한다. 목표고객에 대한 상품의 추천순위 결정에서 목표고객과 이웃사용자의 연령속성을 반영한 추천순위 결정방법을 고안한다. 기존에 발표한 연구인 협력적 필터링 추천 시스템의 정확도 향상의 추천 시스템에서 상품의 추천순위는 상품의 예측 선호도로 결정한다[2].

그러나 선호도는 목표고객의 속성에 따라서 달라질 수 있기 때문에 목표고객과 이웃사용자의 속성을 반영

† 이 논문은 인하대학교 교내 연구비 지원에 의해 연구되었음.

† 교신저자: 이석환, 인천시 남구 용현동 253, 인하대학교 산업공학과 2북 670A

M · P: 010-5514-6242, E-mail: shleekun@inha.ac.kr

2011년 3월 15일 접수; 2011년 6월 1일 수정본 접수; 2011년 6월 7일 게재확정

하지 않고 추천순위를 결정하게 되면 추천의 정확도가 낮아질 수 있다. 이 문제의 해결을 위해 추천순위를 결정할 때 예측 선호도 외에도 목표고객과 이웃사용자의 연령속성의 유사도를 반영하여 추천순위를 결정한다.

다음으로 본 연구에서 제안한 방법을 실제로 고객의 선호도를 조사한 MovieLens 데이터에 적용하고 그 결과를 본 연구와 동일한 MovieLens 데이터를 사용한 기존 연구의 추천 실험결과와 비교한다[2][8].

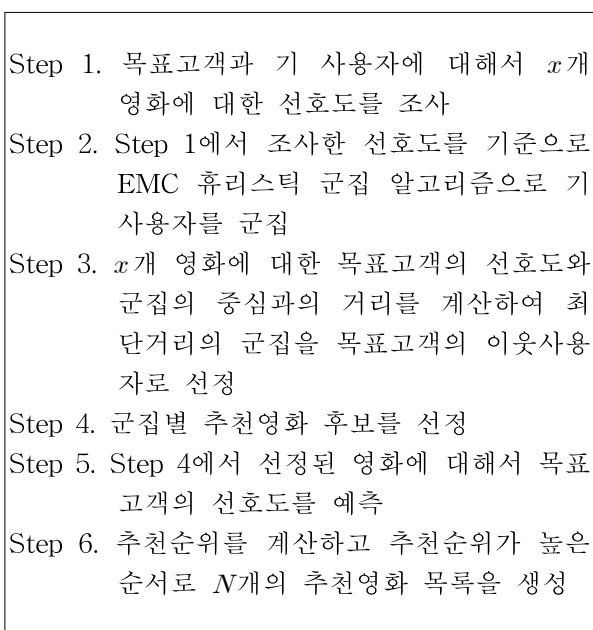
2. 추천 시스템

협력적 필터링 추천 시스템의 성능 향상을 위해 본 연구에서 제시한 방법은 다음과 같은 순서로 설명한다.

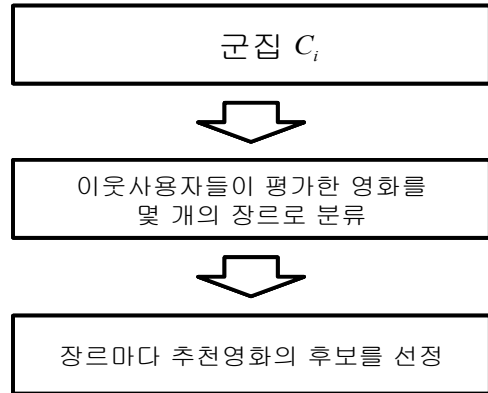
2.1에서는 기존에 발표한 연구인 협력적 필터링 추천 시스템의 정확도 향상을 중심으로 추천 시스템의 추천 과정을 설명하고 2.2에서는 추천 시스템의 정확도를 향상시키기 위해 본 연구에서 제시한 연령속성을 반영한 추천순위 결정방법에 대해 설명한다[2].

2.1 추천과정

본 연구에서 사용한 추천 시스템은 협력적 필터링 추천 시스템의 정확도 향상에서 구축한 시스템과 동일하다. 따라서 추천과정은 기존의 연구와 동일하기 때문에 간략히 설명하기로 한다. <그림 2-1>은 그 추천과정으로 각 단계별 자세한 내용은 <그림 2-1>의 순서로 설명한다[2].



<그림 2-1> 추천과정



<그림 2-2> 장르별 추천영화 후보 선정

기 사용자 군집 및 이웃사용자 선정은 <그림 2-1>에서 Step 1~3에 해당한다. Step 1과 Step 2에서는 기 사용자를 군집하고 Step 3에서는 목표고객의 이웃사용자를 선정한다.

추천영화의 후보선정은 목표고객의 이웃사용자가 평가한 영화를 대상으로 이루어진다. 먼저 이웃사용자가 평가한 영화에 대해서 장르를 구분한 후 각 장르별로 몇 편의 영화를 추천영화의 후보로 선정한다[3]. 여기에서 추천영화 후보는 이웃사용자들의 선호도 합이 높은 영화를 우선적으로 선정한다. <그림 2-2>는 특정 군집에 대해서 추천영화 후보를 선정하는 과정으로 <그림 2-1>의 Step 4에 해당한다.

선호도 예측은 <그림 2-1>에서 Step 5로 앞 단계인 Step 4에서 선정한 추천영화 후보에 대해서 목표고객에 대한 선호도를 예측하는 단계이다. 일반적으로 영화 추천 시스템은 목표고객이 특정 영화에 대해서 갖는 선호도를 예측하고 그 중에서 예측 선호도가 가장 높은 영화를 추천한다. 본 연구에서 사용한 선호도 예측은 식 (1)과 같다.

$$R_{A,m} = \frac{\sum_{j=1}^c R_{j,m}}{c} + \frac{\sum_{j=1}^c w(A,j)(R_{j,m} - \bar{R}_j)}{\sum_{j=1}^c |w(A,j)|} \quad (1)$$

- A : 목표고객
- j : 이웃사용자
- \bar{R}_j : 이웃사용자 j 의 평균 선호도
- c : 이웃사용자의 수
- m : 영화
- $w(A,j)$: 목표고객 A 와 이웃사용자 j 의 유사도 가중치
- $R_{j,m}$: m 영화에 대한 이웃사용자 j 의 선호도

식 (1)에서 $w(A,j)$ 는 피어슨 상관계수이다[4][7]. 상관관계가 너무 낮은 이웃사용자들의 선호도 정보를 반

영할 경우 선호도 예측의 정확도가 낮아질 수 있다. 따라서 본 연구에서는 선호도 예측의 정확도를 높이기 위해 0.5 이상의 상관관계를 갖는 이웃사용자에 대해서만 선호도 가중치를 적용하였다.

한편 목표고객과 부의 상관을 갖는 이웃사용자는 목표고객의 이웃사용자라고 할 수 없다. 따라서 목표고객과 부의 상관을 갖는 이웃사용자는 선호도 예측에서 제외하였다.

2.2 연령속성을 반영한 추천순위 결정방법

추천순위의 결정은 <그림 2-1>에서 Step 6에 해당한다. 추천 시스템에서는 예측 선호도가 높은 순서로 상위 N 개의 영화를 목표고객에게 추천한다. 여기에서 예측 선호도는 목표고객과 이웃사용자의 속성이 같을 경우 비교적 정확하다고 할 수 있다. 따라서 top- N 추천방법에서 N 개의 영화를 선정할 때 목표고객과 이웃사용자의 속성이 유사한 정도에 따라서 예측 선호도에 가중치를 부여하고 이것을 추천순위로 사용하면 추천의 정확도를 향상시킬 수 있다. 추천에 사용자의 속성을 이용한 연구는 ‘개인화 추천 시스템에서 속성 정보를 이용한 연관 사용자 군집 방법’이 있다[5]. 이 연구는 목표고객의 이웃사용자를 선정하는 과정에서 사용자 성별과 연령을 고려하였다. 본 연구에서는 추천 영화 후보에서 N 개의 영화를 선정하기 위해 목표고객과 이웃사용자의 연령속성을 고려한 추천순위를 이용한다. 기존의 top- N 추천방법에서는 예측 선호도만으로 N 개의 영화를 선정하였으나 본 연구에서는 추천순위라는 기준을 사용하여 영화를 선정한다. 추천순위는 예측 선호도에 연령속성을 결합한 것으로 그 연령속성은 연령차와 연령분산이다.

먼저 연령차에 대해 설명한다. 일반적으로 연령대 별로 선호하는 영화가 존재한다. 예를 들면, M1이라는 영화를 30대가 주로 보았다면 M1 영화는 다른 30대가 선호할 확률이 높아진다. 즉, 목표고객의 연령과 M1 영화를 평가한 이웃사용자들의 평균연령이 비슷하다면 M1 영화는 목표고객이 선호할 확률이 높아진다고 할 수 있다.

다음은 연령분산으로 M1이라는 영화를 평가한 사람들의 연령분산이 크다면 그 영화는 다양한 계층에서 보는 영화라고 할 수 있다. 일반적으로 다양한 계층에서 선호하는 영화는 모든 연령층에서 좋은 평가를 받을 수 있는 확률이 크다. 그러므로 이와 같은 영화는 목표고객으로부터 좋은 평가를 받게 될 확률이 커진다.

따라서 추천영화 후보의 예측 선호도에 연령차와 연

령분산을 가중치로 부여하여 그 값이 큰 영화를 우선으로 추천하면 추천의 정확도를 향상시킬 수 있다. 연령차와 연령분산의 가중치를 계산하기 위해서는 먼저 연령차와 연령분산을 계산해야한다. 식 (2)는 연령차 $AgeD_{A,m}$ 을 계산하는 식이고 식 (3)은 연령분산 $AgeV_{A,m}$ 을 계산하는 식이다. 식 (2)에서 Age_A 는 목표고객 A 의 연령이고 $Age_{j,m}$ 는 m 영화를 평가한 이웃사용자의 j 연령이다. 식 (3)에서 $Age_{i,m}$ 는 m 영화를 평가한 이웃사용자 i 의 연령이다.

$$AgeD_{A,m} = \left| Age_A - \frac{\sum_{j=1}^c Age_{j,m}}{c} \right| \quad (2)$$

Age_A : 목표고객 A 의 연령
 $Age_{j,m}$: m 영화를 평가한 이웃사용자 j 의 연령
 c : 이웃사용자의 수

$$AgeV_{A,m} = \sum_{i=1}^c \left(Age_{i,m} - \frac{\sum_{j=1}^c Age_{j,m}}{c} \right)^2 \quad (3)$$

$Age_{i,m}$: m 영화를 평가한 이웃사용자 i 의 연령
 c : 이웃사용자의 수

연령차와 연령분산은 예측 선호도와 동일하게 5점 척도로 변환한다. 연령차의 경우는 목표고객과 이웃사용자의 연령차가 작을수록 목표고객과 이웃사용자간의 유사도가 높다고 할 수 있다. 반대로 연령차가 커지면 유사도가 낮다고 할 수 있다. 따라서 연령차가 작으면 연령차 가중치에는 큰 값을 부여하고 연령차가 크면 연령차 가중치에는 작은 값을 부여해야 한다. 식 (4)에서 kd_A 는 연령차 $AgeD_{A,m}$ 을 5점 척도로 변환하는 계수이다. kd_A 는 식 (5)에서 연령차 가중치 $ADR_{A,m}$ 을 계산하기 위해 사용한다.

$$kd_A = \frac{5}{\max(AgeD_{A,m})} \quad (4)$$

$$ADR_{A,m} = 5 - kd_A AgeD_{A,m} \quad (5)$$

연령분산은 연령차와는 반대로 값이 작을수록 목표고객의 선호도가 낮고 값이 커질수록 선호도가 높다고

할 수 있다. 따라서 연령분산이 작으면 연령분산 가중치에는 작은 값을 부여하고 연령분산이 크면 연령분산 가중치에는 큰 값을 부여해야 한다. 식 (6)에서 kv_A 는 연령분산 $Age V_{A,m}$ 을 5점 척도로 변환하는 계수이다.

kv_A 는 식 (7)에서 연령분산 가중치 $AVR_{A,m}$ 을 계산하기 위해 사용한다.

$$kv_A = \frac{5}{\max(Age V_{A,m})} \tag{6}$$

$$AVR_{A,m} = kv_A Age V_{A,m} \tag{7}$$

<표 2-1>과 <표 2-2>는 연령차와 연령분산 가중치를 계산한 예이다. <표 2-1>에서 연령차 계수 kd_A 는 $5/12.5=0.4$ 이다. 따라서 연령차가 2인 경우 연령차 가중치는 $5-(2 \times 0.4)=4.2$ 가 된다. <표 2-2>에서 연령분산 계수 kv_A 는 $5/15.57=0.32$ 이다. 따라서 연령분산이 10인 경우 연령분산 가중치는 $10 \times 0.32=3.2$ 가 된다.

연령차와 연령분산의 가중치가 계산되면 이 값은 예측 선호도에 더해져서 추천순위를 결정할 수 있다. 식 (8)은 추천순위 점수 $RR_{A,m}$ 를 계산하는 식으로 예측 선호도 $R_{A,m}$ 에 연령차 가중치 $ADR_{A,m}$, 연령분산 가중치 $AVR_{A,m}$ 을 합한 것이다. α 와 β 는 연령차 가중치와 연령분산 가중치의 적용비율을 정하기 위한 계수이다. 계수의 값을 알아내기 위한 실험에서 $\alpha=0.5$, $\beta=0.5$ 일 경우 추천의 정확도가 가장 높다는 것을 확인하였다. α 와 β 를 결정하기 위한 실험은 3.2.1에서 설명한다.

<표 2-1> 연령차 가중치 계산

연령차	연령차 가중치
0	5
2	4.2
:	:
12.5	0

<표 2-2> 연령분산 가중치 계산

연령분산	연령분산 가중치
15.57	5
:	:
10	3.2
0	0

$$RR_{A,m} = R_{A,m}(1 - (\alpha + \beta)) + ADR_{A,m}\alpha + AVR_{A,m}\beta \tag{8}$$

$R_{A,m}$: m 영화에 대한 목표고객 A 의 예측 선호도
 α : 연령차 가중치 적용비율
 β : 연령분산 가중치 적용비율
 $ADR_{A,m}$: m 영화에 대한 목표고객 A 의 연령차 가중치
 $AVR_{A,m}$: m 영화에 대한 목표고객 A 의 연령분산 가중치

3. 추천 시스템 성능평가

본 연구에서 제시한 연령속성을 반영한 추천순위 결정방법(2.1 참조)을 추천 시스템에 적용함으로써 기존의 추천 시스템보다 추천의 정확도가 향상된다는 것을 실험을 통하여 확인한다. 먼저 3.1에서는 추천실험의 과정을 설명한다. 3.2에서는 본 연구에서 제시한 연령속성을 반영한 추천순위 결정방법을 사용했을 경우와 기존의 예측 선호도만으로 추천순위를 결정하였을 경우 선호도 예측의 정확도를 비교한다.

3.1 추천실험 과정

본 연구에서 제안한 추천 시스템의 성능평가를 위해 <그림 3-1>과 같은 절차로 추천실험을 진행하였다. 추천실험은 3회 반복하여 진행하였다. 추천실험과정은 기존의 연구와 동일하기 때문에 간략히 설명한다[2].

Step 1은 전체 사용자를 train 집단과 test 집단으로 나누는 과정으로 test 집단은 목표고객이고 train 집단은 기 사용자이다. train 집단은 군집과정을 통해서 목표고객의 이웃사용자로 된다.

- | |
|---|
| Step 1. 전체 사용자를 train 사용자와 test 사용자로 나눔 |
| Step 2. 군집의 기준으로 사용할 영화 10개를 랜덤으로 선택 |
| Step 3. Step 2에서 선택한 10개 영화의 선호도로 train 사용자를 군집 |
| Step 4. test 사용자의 선호도와 Step 3에서 생성한 군집 사이의 거리를 계산하여 최단 거리를 갖는 군집을 test 사용자의 이웃사용자로 결정 |
| Step 5. 군집별 추천영화의 후보를 선정 |
| Step 6. Step 5에서 선정된 영화에 대해서 test 사용자의 선호도를 예측 |
| Step 7. test 사용자별 추천영화 후보에 대해서 추천순위를 계산하고 추천순위가 높은 순서로 10개의 영화를 추천 |

<그림 3-1> 추천 실험 절차

Step 2는 train 집단을 군집하기 위해 군집의 기준으로 사용할 영화를 선택하는 단계이다.

Step 3은 train 사용자를 군집하는 단계로 군집 알고리즘은 EMC 휴리스틱 군집 알고리즘을 사용하였다[1].

군집의 수는 5개부터 15개까지 증가시키면서 수행한다.

Step 4는 목표고객인 test 사용자의 이웃사용자를 선정하는 단계이다.

Step 5는 목표고객인 test 사용자에게 추천할 영화의 후보들을 선정하는 단계이다. 추천영화 후보는 <그림 2-2>와 같이 이웃사용자가 평가한 영화중에서 선호도 합이 큰 영화를 장르별로 3편 선정하였다. MovieLens 는 19개의 장르가 존재하므로 총 57개의 추천영화 후보가 선정된다.

Step 6은 Step 5에서 선정한 추천영화 후보에 대해서 test 사용자들의 선호도를 예측하는 단계로 이웃사용자 중심 선호도 예측방법을 사용한다[2].

Step 7은 Step 6에서 계산된 예측 선호도로 추천순위를 정하는 단계이다. 추천순위는 본 연구에서 제시한 연령속성을 반영한 추천순위결정 방법으로 최종 추천 목록을 생성한다.

추천목록이 생성되면 추천 시스템의 성능을 평가하기 위해 추천목록의 영화에 대한 예측 선호도와 test 사용자가 입력한 실제 선호도를 비교하여 추천의 정확도를 평가한다.

본 연구에서 제시한 추천 시스템의 성능은 MAE (Mean Absolute Error)를 사용하여 평가한다. 추천 시스템의 성능은 MAE 값이 작을수록 우수하다고 평가한다[4][6][7]. MAE는 식 (9)와 같다.

$$MAE = \frac{\sum_{m=1}^n |P_{A_m} - E_{A_m}|}{n} \tag{9}$$

P_{A_m} : 목표고객 A의 실제 선호도

E_{A_m} : 목표고객 A의 예측 선호도

n : 추천영화중에서 목표고객 A가 실제로 관람한 영화편수

P_{A_m} 는 목표고객 A가 m 영화에 대해 평가한 실제 선호도이고 E_{A_m} 는 추천 시스템에 의해서 예측된 선호도이다. n 은 목표고객인 test 사용자에게 추천한 영화중에서 test 사용자가 실제로 평가한 영화의 개수이다.

3.2 실험결과 및 분석

본 연구에서는 추천의 정확도를 향상시키기 위해 추천목록을 보다 정확하게 생성할 수 있는 연령속성을 반영한 추천순위 결정방법을 제시하였다. 여기서는 기존 연구인 협력적 필터링 추천 시스템의 정확도 향상의 추천 시스템에 연령속성을 반영한 추천순위결정방법을 적용할 경우 추천의 정확도가 향상된다는 것을 실험을 통하여 확인한다[2]. 이 실험은 <그림 3-1>의 추천실험 절차에서 마지막 단계인 Step 7에 해당한다.

3.2.1에서는 연령차 가중치와 연령분산 가중치의 적용비율을 결정하는 방법을 설명한다. 3.2.2에서는 3.2.1에서 결정한 연령차 가중치와 연령분산 가중치의 적용비율을 추천 시스템에 적용하여 추천의 정확도가 향상되는 것을 확인한다.

<표 3-1> 예측 선호도와 연령차 가중치를 적용한 추천결과

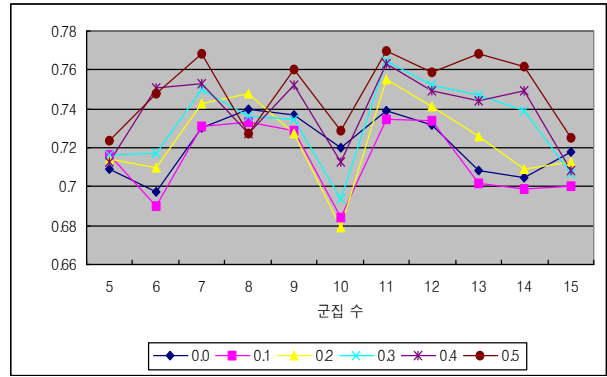
α 군집	0.00	0.10	0.20	0.30	0.40	0.50
5	0.709	0.716	0.714	0.716	0.713	0.724
6	0.697	0.690	0.710	0.717	0.751	0.748
7	0.730	0.731	0.743	0.750	0.753	0.768
8	0.740	0.733	0.748	0.736	0.727	0.727
9	0.737	0.729	0.727	0.735	0.752	0.760
10	0.720	0.684	0.679	0.694	0.713	0.729
11	0.739	0.735	0.755	0.765	0.763	0.770
12	0.732	0.734	0.741	0.752	0.749	0.759
13	0.708	0.702	0.726	0.747	0.744	0.768
14	0.705	0.699	0.709	0.739	0.749	0.762
15	0.718	0.700	0.713	0.707	0.708	0.725
평균	0.721	0.714	0.724	0.733	0.738	0.749

3.2.1 연령차 가중치와 연령분산 가중치의 적용 비율

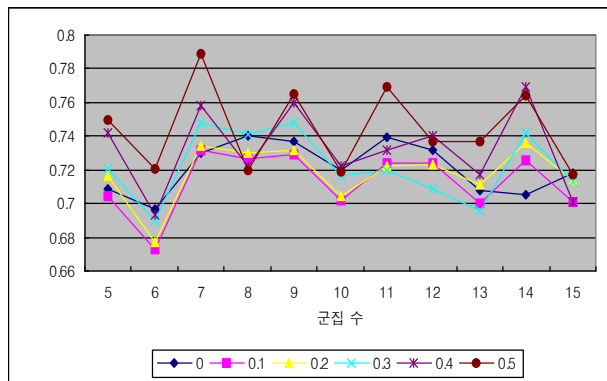
추천순위 결정에 사용하는 식에는 연령차 가중치와 연령분산 가중치 및 그 값에 해당하는 적용비율 α 와 β 가 있다. 먼저 이들 적용비율 α 와 β 값을 결정하는 실험을 하였다. 연령차 가중치 적용비율 α 를 결정하는 실험에서 α 는 0부터 0.5까지 0.1씩 증가시키면서 MAE를 계산하였다. <표 3-1>은 연령차 가중치에 적용비율 α 를 곱하여 예측 선호도에 더한 후 그 값이 큰 순서로 영화를 추천했을 때 MAE를 표시한 것으로 음영부분은 군집 별로 가장 좋은 MAE를 나타낸 것이다. MAE 변화의 경향을 보기 위해 <표 3-1>의 결과를 <그림 3-2>와 같이 나타내었다. <그림 3-2>와 같이 MAE는 α 가 0.1일 때 모든 군집 수에 대해서 대체로 좋은 값을 나타내었다.

연령분산 가중치 적용비율 β 를 결정하는 실험도 α 와 동일하게 β 를 0부터 0.5까지 0.1씩 증가시키면서 MAE를 계산하였다. <표 3-2>는 연령분산 가중치에 적용비율 β 를 곱하여 예측 선호도에 더한 후 그 값이 큰 순서로 영화를 추천했을 경우 MAE를 표시한 것으로 음영부분은 군집 별로 가장 좋은 MAE를 나타낸다.

MAE 변화의 경향을 보기 위해 <표 3-2>의 결과를 <그림 3-3>과 같이 나타내었다. <그림 3-3>과 같이 MAE는 α 의 경우와 동일하게 β 가 0.1일 때 모든 군집 수에 대해서 대체로 좋은 값을 나타내었다.



<그림 3-2> 예측 선호도와 연령차 가중치를 적용한 추천결과



<그림 3-3> 예측 선호도와 연령분산 가중치를 적용한 추천결과

<표 3-2> 예측 선호도와 연령분산 가중치를 적용한 추천결과

β 군집	0.00	0.10	0.20	0.30	0.40	0.50
5	0.709	0.704	0.716	0.721	0.742	0.750
6	0.697	0.673	0.677	0.690	0.693	0.721
7	0.730	0.732	0.734	0.748	0.758	0.789
8	0.740	0.727	0.730	0.741	0.722	0.720
9	0.737	0.729	0.732	0.748	0.760	0.765
10	0.720	0.702	0.704	0.716	0.722	0.719
11	0.739	0.724	0.722	0.720	0.732	0.769
12	0.732	0.724	0.723	0.709	0.740	0.737
13	0.708	0.700	0.711	0.696	0.717	0.737
14	0.705	0.726	0.736	0.742	0.769	0.764
15	0.718	0.701	0.714	0.713	0.701	0.717
평균	0.721	0.713	0.718	0.722	0.732	0.744

실험결과 연령차 가중치 적용비율 α 와 연령분산 가중치의 적용비율 β 는 각각 0.1이 적합하다는 것을 확인하였다.

다음은 추천순위 결정에 앞에서 결정한 연령차 가중

치와 연령분산 가중치를 동시에 적용하는 실험을 하였다. 연령차 가중치 또는 연령분산 가중치만을 적용하였을 때 α 와 β 는 각각 0.1이었으므로 두 가지 모두를 적용시키기 위해 α 와 β 를 각각 0.05로 설정하였다.

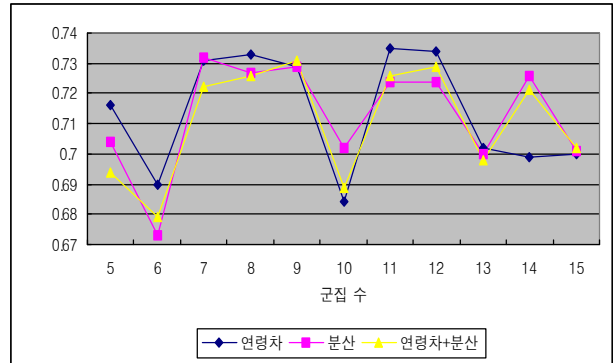
<표 3-3> 연령속성을 반영한 MAE 비교

군집 수	연령차 $\alpha=0.1$	연령분산 $\beta=0.1$	연령차와 연령분산의 모두 적용 $\alpha=0.05, \beta=0.05$
5	0.716	0.704	0.694
6	0.690	0.673	0.679
7	0.731	0.732	0.722
8	0.733	0.727	0.726
9	0.729	0.729	0.731
10	0.684	0.702	0.689
11	0.735	0.724	0.726
12	0.734	0.724	0.729
13	0.702	0.700	0.698
14	0.699	0.726	0.721
15	0.700	0.701	0.702
평균	0.714	0.713	0.711

<표 3-3>은 추천순위의 결정에 연령차 가중치만을 적용한 경우, 연령분산 가중치만을 적용한 경우와 연령차와 연령분산의 가중치 모두를 적용한 경우의 MAE를 표시한 것이다. 음영부분은 세 가지 방법 중에서 MAE가 가장 좋은 것을 나타낸 것이다. 연령차 가중치만을 적용한 경우, 연령분산 가중치만을 적용한 경우, 연령차와 연령분산의 가중치 모두를 적용한 경우 전체 실험횟수에서 각각 4회씩 우세하였다. MAE가 우세한 횟수는 세 가지 경우 모두 같기 때문에 어떤 방법이 우수한 방법인지 알 수 없다. 이 문제를 해결하기 위해 모든 군집 수에 대한 평균 MAE를 계산하였다. 그 결과 연령차 가중치만을 적용한 경우는 0.714, 연령분산 가중치만을 적용한 경우는 0.713, 연령차와 연령분산의 가중치 모두를 적용한 경우는 0.711이었다. 따라서 추천순위에 연령차와 연령분산의 가중치 모두 적용시킬 경우 MAE가 향상된다는 것을 확인할 수 있다.

MAE 변화의 경향을 보기 위해 <표 3-3>의 결과를 <그림 3-4>와 같이 그래프로 나타내었다. 군집 수에 따른 MAE의 변화는 군집 수 14를 제외하고 세 가지 경우 모두 유사한 경향을 나타내었다.

위의 실험 이외에 α 와 β 를 각각 2.0까지 증가시키면서 실험을 하였으나 향상 폭은 점차 낮아졌다. 이 결과로 추천순위 결정에서 연령차와 연령분산은 예측 선호도를 보완할 수 있다는 것을 확인할 수 있다.



<그림 3-4> 연령속성을 반영한 추천결과 비교

3.2.2 연령속성을 반영한 추천순위 결정

3.2.1의 결과를 바탕으로 다음과 같은 실험을 진행하였다. 이 실험은 이웃사용자 중심 선호도 예측방법으로 선호도를 예측하고 앞의 실험(3.2.1 참조)에서 결정된 α 와 β 를 추천순위 결정 식에 적용하는 것이다. <표 3-4>는 추천순위 결정에 예측 선호도만을 사용한 경우와 연령차와 연령분산의 가중치를 적용한 경우의 MAE를 비교한 표이다. 음영부분은 두 가지 방법 중에서 가장 좋은 MAE를 나타낸 것이다. 1회 차 실험에서 MAE가 크게 향상된 군집 수는 10으로 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 4.31% 향상되었다. 반면 군집 수가 14인 경우는 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 향상되지 않았다. 그 외의 군집 수에서는 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 최소 0.41%에서 최대 2.58%까지 향상되었다.

<표 3-4> 연령속성을 반영한 MAE 비교-이웃사용자 중심 선호도 예측 (1회)

군집 수	기존 선호도 예측방법	예측 선호도만을 적용한 추천순위	연령차, 연령분산 가중치를 적용한 추천순위
5	0.788	0.709	0.694
6	0.782	0.697	0.679
7	0.757	0.730	0.722
8	0.762	0.740	0.726
9	0.756	0.737	0.731
10	0.757	0.720	0.689
11	0.760	0.739	0.726
12	0.770	0.732	0.729
13	0.750	0.708	0.698
14	0.799	0.705	0.721
15	0.778	0.718	0.702

<표 3-5>는 2회 차 실험의 결과이다. 음영부분은 두 가지 방법 중에서 가장 좋은 MAE를 나타낸 것이다.

MAE가 크게 향상된 군집 수는 5, 7로 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 각각 2.32%, 2.27% 향상되었다. 반면 군집 수가 6, 13, 14인 경우는 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 향상되지 않았다. 그 외의 군집 수에서는 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 최소 0.42%에서 최대 1.79%까지 향상되었다.

<표 3-5> 연령속성을 반영한 MAE 비교-이웃사용자 중심 선호도 예측- (2회)

군집 수	기존 선호도 예측방법	예측 선호도만을 적용한 추천순위	연령차, 연령분산 가중치를 적용한 추천순위
5	0.771	0.707	0.691
6	0.743	0.707	0.723
7	0.761	0.722	0.706
8	0.787	0.768	0.759
9	0.715	0.680	0.674
10	0.733	0.705	0.694
11	0.751	0.738	0.725
12	0.778	0.747	0.736
13	0.711	0.693	0.708
14	0.741	0.722	0.733
15	0.784	0.721	0.718

<표 3-6> 연령속성을 반영한 MAE 비교-이웃사용자 중심 선호도 예측- (3회)

군집 수	기존 선호도 예측방법	예측 선호도만을 적용한 추천순위	연령차, 연령분산 가중치를 적용한 추천순위
5	0.782	0.792	0.803
6	0.751	0.726	0.724
7	0.763	0.747	0.731
8	0.796	0.722	0.720
9	0.740	0.674	0.696
10	0.735	0.723	0.693
11	0.747	0.744	0.735
12	0.743	0.709	0.699
13	0.728	0.745	0.705
14	0.732	0.735	0.720
15	0.722	0.714	0.696

<표 3-6>는 3회 차 실험의 결과이다. 음영부분은 두 가지 방법 중에서 가장 좋은 MAE를 나타낸 것이다. MAE가 크게 향상된 군집 수는 13으로 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 5.37% 향상되었다. 반면 군집 수가 5, 9인 경우는 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 향상되지 않았다. 그 외의 군집 수에서는 추천순위에 연령차와 연령분산의 가중치를 적용한 경우가 예측 선호도만을 적용한 경우에 비해서 최소 0.28%에서 최대 4.15%까지 향상되었다.

4. 결론

본 연구에서는 협력적 필터링 기법을 적용한 추천 시스템에서 추천의 정확도를 향상시키기 위해 연령속성을 반영한 추천순위 결정방법을 고안하였다. 추천순위 결정방법에서 기존의 예측 선호도만을 고려하였을 경우는 목표고객과 이웃사용자의 속성을 충분히 반영하지 못했기 때문에 추천의 정확도가 더 이상 향상되지 않았다. 본 연구에는 목표고객과 이웃사용자의 속성인 연령차와 연령분산을 추천순위 결정에 반영하였기 때문에 기존 논문의 예측 선호도만으로 추천순위를 결정하는 방법에 비해 추천의 정확도를 향상시킬 수 있었다. 기존 논문의 방법에 본 연구에서 고안한 연령속성을 반영한 추천순위 결정방법을 사용했을 경우 MAE는 기존 논문에 비해서 최대 5.37% 까지 향상되었다.

이와 같이 추천 시스템의 추천 정확도가 향상된 이유는 추천목록의 개수에 제한이 있는 경우에 연령속성을 이용한 새로운 추천순위 결정방법을 사용함으로써 보다 정확한 추천목록을 생성하였기 때문이다. 본 연구에서는 추천순위 결정에 영향을 주는 연령차와 연령분산을 발견하여 추천의 정확도를 향상시켰지만 기존 논문의 추천 정확도에 비해 그 향상 폭은 크지 않았다.

향후 연구과제는 연령속성 이외의 타 속성을 이용하여 추천의 정확도를 더욱 향상시키는 것이다.

5. 참고 문헌

[1] 이석환, 박승현, “군집의 효율향상을 위한 휴리스틱 알고리즘”, 대한안전경영과학회, 제11권 제3호, 2009, 157-166
 [2] 이석환, 박승현, “협력적 필터링 추천 시스템의 정확도 향상”, 대한안전경영과학회, 제12권 제1호, 2010, 127-136

- [3] 이재식, 박석두, “장르별 협업필터링을 이용한 영화 추천 시스템의 성능 향상”, 한국지능정보시스템학회, 제13권 제4호, 2007, 65-78
- [4] 정경용, 이정현, “개인화 추천 시스템의 예측 정확도 향상을 위한 사용자 유사도 가중치에 대한 비교 평가”, 전자공학회, 제42권 제6호, 2005, 63-73
- [5] 한경수, 조동주, 정경용, “개인화 추천 시스템에서 속성 정보를 이용한 연관 사용자 군집 방법”, 한국정보과학회, 제33권 제2(B)호, 2006, 169-173
- [6] Ahn, H. J. “A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem”, Information Sciences, 178, 2008, 37-51
- [7] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms”, Association for Computing Machinery, 2001, 285-295
- [8] MovieLens dataset, URL : <http://movielens.umn.edu>, 1996

저 자 소 개

이 석 환



인하대학교 산업공학과에서 공학사, 공학석사 및 공학박사 학위를 취득하였다. 주요 관심 분야는 데이터 마이닝이다.

주소: 인천광역시 남구 용현동 253, 인하대학교 산업공학과

박 승 현



인하대학교 금속공학과에서 공학사, 일본 Keio대학 관리공학과에서 공학석사 및 공학박사 학위를 취득 하였다. 현재 인하대학교 산업공학과 교수로 재직 중이다. 주요 관심 분야는 FMS와 각종 생산시스템의 설계 및 운영, 인터넷 마케팅과 데이터 마이닝 등이다.

주소: 인천광역시 남구 용현동 253, 인하대학교 산업공학과