

협업 필터링 추천 시스템을 위한 데이터 신뢰도 기반 가중치를 이용한 하이브리드 선호도 예측 기법

이 오 준*, 백 영 태**

Hybrid Preference Prediction Technique Using Weighting based Data Reliability for Collaborative Filtering Recommendation System

O-Joun Lee *, Yeong-tae Baek **

요 약

협업 필터링 추천은 사용자의 아이템에 대한 선호도를 기반으로 유사 아이템 집합 또는 유사 사용자 집합을 생성하고 이를 이용해 사용자의 특정 아이템에 대한 선호도를 예측한다. 따라서 선호도 행렬이 희박할 경우, 추천의 신뢰도는 급격히 낮아진다. 본 논문에서는 위 문제를 해결하기 위해 데이터 신뢰도 기반 가중치를 이용한 하이브리드 선호도 예측 기법을 제안한다. 선호도 예측은 유사 아이템 집합과 유사 사용자 집합을 모두 생성하고 각 집합을 통해 사용자의 선호도를 예측하며, 모델의 상황을 반영한 가중치를 이용해 각 예측치를 병합하여 수행된다. 이 기법은 사용자 선호도 예측 정확도를 높이며 선호도 행렬 희박도가 높은 상황에도 추천 서비스의 신뢰도를 유지할 수 있도록 한다. 이 기법을 바탕으로 추천 시스템을 구현하고 절대평균오차를 기준으로 서비스 신뢰도 향상을 측정하였다. 실험에서 본 기법은 Hao Ji가 제안한 기존의 기법에 비해 선호도 행렬 희박도가 84% 이상인 상황에서 평균 21.7%의 성능 향상을 보여 효과적으로 행렬 희박도 문제를 해소할 수 있음을 검증하였다.

▶ Keywords : 추천, 협업 필터링, 선호도 예측

Abstract

Collaborative filtering recommendation creates similar item subset or similar user subset based on user preference about items and predict user preference to particular item by using them. Thus, if preference matrix has low density, reliability of recommendation will be sharply decreased. To

•제1저자 : 이오준 •교신저자 : 백영태

•투고일 : 2014. 4. 1, 심사일 : 2014. 4. 18, 게재확정일 : 2014. 5. 3.

* 단국대학교 소프트웨어학과(Dept. of Software science, Dankook University)

** 김포대학교 멀티미디어과(Dept. of Multimedia, Kimpo University)

solve these problems we suggest Hybrid Preference Prediction Technique Using Weighting based Data Reliability. Preference prediction is carried out by creating similar item subset and similar user subset and predicting user preference by each subset and merging each predictive value by weighting point applying model condition. According to this technique, we can increase accuracy of user preference prediction and implement recommendation system which can provide highly reliable recommendation when density of preference matrix is low. Efficiency of this system is verified by Mean Absolute Error. Proposed technique shows average 21.7% improvement than Hao Ji's technique when preference matrix sparsity is more than 84% through experiment.

▶ Keywords : Recommendation, Collaborative Filtering, Preference Prediction

I. 서 론

추천 시스템을 구축하기 위한 여러 기법 중 널리 사용되는 기법으로 협업 필터링 추천 기법이 있다. 협업 필터링 추천 기법은 다른 사용자의 선호도 정보를 기반으로 아이템을 추천하는 기법이다. 따라서 해당 특정 도메인에 대한 지식을 필요로 하지 않으며, 사용자와 아이템에 대한 광범위한 데이터를 필요로 하지도 않는다. 또한, 사용자 프로파일과 아이템 특성 정보의 정확도의 한계로 인해 발생하는 문제로부터 자유롭다. 하지만, 초기 평가 문제, 초기 사용자 문제, 확장성 문제, 행렬 희박성 문제와 같은 근본적인 한계점 또한 갖고 있다[1,2,3,4]. 이 문제들을 해결하기 위해 아래와 같은 다양한 기법들이 제안되었으나, 이 문제들을 모두 해결하지는 못했다.

Y. Song은 군집화 기법을 적용하여 사용자 선호도 예측의 효율성을 높이는 방법을 제안하였다[2]. 이 방법은 선호도 예측의 시간 효율성을 높일 수 있었지만, 기존의 방법에 비해 선호도 예측의 신뢰도를 개선하지는 못하였다.

J. Read는 토큰화와 최소분류오차(MCE)를 이용하여 음악을 내용 기반으로 분석하여 음악에 대한 사용자 선호도를 예측하는 방법을 제안하였다[3]. 이 방법은 기존 방법들에 비해 개선된 성능을 보이지만, 다른 종류의 아이템들에 적용하기 힘들다는 문제점을 갖는다.

H. Ji는 행렬 희박성 문제를 개선하기 위해 사용자 선호도 예측 시, 사용자 기반 협업 필터링과 아이템 기반 협업 필터링을 병용하는 방법을 제안하고 실험을 통해 두 예측 값을 결합하기 위한 가중치를 결정하였다[4]. 이 방법은 사용자 선호

도 예측의 신뢰도를 향상 시킬 수 있지만, 가중치 값이 추천 모델의 상태를 반영하지 못한다는 문제점을 갖는다.

본 연구에서는 위와 같은 문제점들을 개선하기 위해 두 가지 방법을 이용한다. 첫째, 사용자 기반 협업 필터링 기법과 아이템 기반 협업 필터링 기법을 병용하며, 둘째, 모델의 상황을 반영한 가중치를 이용해 두 예측 값을 결합한다. 또한, 위 방법들을 통해 행렬 희박성 문제를 개선한 새로운 형태의 협업 필터링 추천 시스템을 설계하고 웹과 모바일 상에서 이용 가능하도록 구현한다.

이를 위해 2장에서는 협업 필터링 추천 기법에 관한 기존 연구들을 살펴보고 3장에서는 본 연구에서 제안하는 협업 필터링 추천 기법의 사용자 선호도 예측 알고리즘에 대해 상세히 기술한다. 4장에서는 설계를 바탕으로 구현된 시스템에 대해 검증하고 평가하며 5장에서는 결론 및 향후 연구 방향을 제시하고자 한다.

II. 관련 연구 및 배경 지식

1. 협업 필터링 추천 시스템

협업 필터링 추천 기법은 사용자들의 선호도 정보를 기반으로 새로운 사용자가 관심을 가질 것으로 예측되는 항목을 추천해주는 기법이다[5,6,7,8]. 규칙 기반 필터링 추천 기법이나 내용 기반 필터링 추천 기법 등의 방법이 항목 자체의 속성 정보를 이용해서 사용자에게 아이템을 추천하는 것과는 달리 협업 필터링 추천 기법은 아이템에 대한 다른 사용자들의 선호를 기반으로 한다. 한편, 이전의 추천 시스템들이 대

부분 텍스트 기반의 자료를 대상으로 하였으나 협업 필터링 추천 기법은 다양한 멀티미디어 아이템을 추천하기 위한 많은 시도들에 이용되었다. 이는 아이템의 특성을 기반으로 하는 여타 추천 기법들을 멀티미디어에 적용하기에는 멀티미디어 콘텐츠의 정성적인 특성을 정량적으로 수치화하거나, 추출하는데 어려움이 있기 때문이다.

이러한 협업 필터링 추천 기법의 장점은 다음과 같다. 첫째, 협업 필터링 추천 기법을 사용해서 항목의 선호도를 예측할 때 예측 대상이 되는 아이템이 다양한 분야에 속해있다고 가정하면, 사용자 A가 선호하는 분야와 같은 분야의 아이템을 선호했던 이웃이 새로운 분야를 선호하였을 경우 특정 사용자에게 새로운 분야의 아이템을 추천할 수 있다. 둘째, 협업 필터링 추천 기법은 동질적인 선호도를 가진 사용자 집단에서 비교적 정확한 예측이 가능하다. 셋째, 협업 필터링 추천 기법은 많은 계산량을 요구하지 않는 중소규모의 환경에서 실시간 추천을 위해서 고안되었기 때문에 개인화된 추천을 위한 계산은 복잡하지 않고 속도도 빠르다.

그러나 협업 필터링 추천 기법은 초기 평가, 초기 사용자 문제, 확장성 문제, 행렬의 희박성 문제 등의 제약점을 갖고 있다[9,10,11]. 첫째, 초기 평가, 초기 사용자 문제는 평가를 전혀 하지 않은 새로운 사용자에게 아이템을 추천하거나 전혀 평가되지 않은 아이템을 추천하는 것이 어렵다는 것이다. 대안으로 충분한 선호도 정보가 쌓일 때까지 사용자 프로파일의 속성 정보나 아이템의 특성 정보를 이용한 내용 기반 추천 기법을 병행하는 방법들이 연구 되어왔다[12].

둘째로, 확장성 문제는 협업 필터링 추천 기법에서 사용자와 아이템의 수가 증가할수록 유사 사용자 혹은 유사 아이템 집합을 생성하는데 필요한 연산 비용도 비례하여 증가한다는 것이다. 이는 실시간 추천을 어렵게 만드는 요인이 되며, 이를 개선하기 위한 방법 중 하나가 군집화 기법을 적용하는 것이다[12].

셋째로, 행렬의 희박성 문제는 일반적으로 고객들은 상품 평가를 잘 하지 않는 경향이 있으며, 선호도 행렬이 희박할 경우 추천의 신뢰도를 떨어뜨릴 수 있다는 것이다. 특히 행렬이 희박할 경우, 몇몇 사용자들에 의해서만 높게 평가된 아이템들에 대해서는 유사한 아이템을 찾기 힘들기 때문에 추천이 어려울 수 있다[13]. 희박성 문제를 개선하기 위해 Pazzani는 음식점 추천 시스템에서 성별, 나이, 지역 코드, 학교 직장 등의 인적 정보를 사용하였으며[13], Huang은 사용자간 트랜지티브 연관성을 사용자 탐색 방법에 이용하는 방법을 제안하였다[9]. 또한, Hao Ji는 사용자 기반 선호도 예측 기법과 아이템 기반 선호도 예측 기법을 병용하여 선호도 예측의 정

확도를 높이는 방법을 제안하였다[4].

2. 아이템 기반 선호도 예측 기법

아이템 기반 선호도 예측에서 이웃 관계는 아이템 간의 관계에 의해 구성된다[14,15,16]. 아이템들의 부분 집합은 특정 아이템과 다른 아이템들 사이의 유사도를 바탕으로 선택된다. 아이템들에 대한 선호도 가중치 조합은 누락된 선호도를 추정하는데 이용된다. 표 1은 아이템 기반 선호도 예측 모델의 전체 알고리즘에 대한 요약이다.

표 1. 아이템 기반 선호도 예측 알고리즘
Table 1. Hybrid Preference Prediction Algorithm

아이템 기반 선호도 예측 알고리즘	
item_i, item_j:	i번째와 j번째의 임의의 아이템
item_m:	추천의 대상이 되는 아이템
<pre> for(int i=0; i<num_of_item; i++){ for(int j=0; j<num_of_item; j++){ Calculate_Item_Similarity(item_i, item_j); //(수식 1) } } Make_Similarity_Matrix_of_Item(); Choose_num_of_neighborhood_Item(k, item_m); Predict_Missing_Rating_about_Item(item_m); //(수식 2) </pre>	

첫 번째 단계에서 아이템 i 와 아이템 j 의 유사도 가중치인 $w_{i,j}$ 는 다음과 같이 구할 수 있다.

$$w_{i,j} = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad (\text{수식 1})$$

U 가 아이템 i 와 j 를 모두를 평가한 적이 있는 사용자들의 집합일 때, $R_{u,i}$ 는 사용자 u 의 아이템 i 에 대한 선호도이고, $R_{u,j}$ 는 사용자 u 의 아이템 j 에 대한 선호도이다. 또한, \bar{R}_i 는 사용자 집합 U 의 아이템 i 에 대한 선호도의 평균이며 \bar{R}_j 는 사용자 집합 U 의 아이템 j 에 대한 선호도의 평균이다.

두 번째 단계에서는, 만약 사용자 a 의 아이템 m 에 대한 평가가 빠져 있다면 아이템 m 과 가장 유사도가 높은 k 개의 아이템을 유사도 테이블을 바탕으로 선택한다.

세 번째 단계에서는 사용자 a 의 아이템 m 에 대한 선호도

를 아이템 가중치의 평균을 이용해 예측한다.

$$p_{a,m} = \frac{\sum_{k \in K} R_{a,k} w_{m,k}}{\sum_{k \in K} w_{m,k}} \quad (\text{수식 2})$$

K 가 아이템 m 과 가장 유사한 k 개의 아이템들의 집합일 때, $p_{a,m}$ 은 사용자 a 의 아이템 m 에 대한 선호도가 된다.

아이템 기반 선호도 예측 기법의 연구사례로, J. Read가 제안한 토큰화와 최소분류오차(MCE)를 이용하여 음악의 내용 기반으로 분석하여 음악에 대한 사용자 선호도를 예측하는 방법이 있다[3].

3. 사용자 기반 선호도 예측 기법

사용자 기반 선호도 예측 알고리즘에서 이웃 관계는 사용자와 사용자의 관계에 의해 구성된다[17,18]. 사용자들의 부분 집합은 특정 사용자와 다른 사용자들 사이의 유사도를 바탕으로 선택된다. 아이템들에 대한 선호도 가중치 조합은 누락된 선호도를 추정하는데 이용된다. 표 2는 사용자 기반 선호도 예측 모델의 전체 알고리즘에 대한 요약이다.

표 2. 사용자 기반 선호도 예측 알고리즘
Table 2. User Based Preference Prediction Algorithm

사용자 기반 선호도 예측 알고리즘
user_i, user_j: i번째와 j번째의 임의의 사용자 user_a: 추천을 요청하는 사용자
<pre> for(int i=0; i<num_of_item; i++){ for(int j=0; j<num_of_item; j++){ Calculate_User_Similarity(user_i, user_j); //(수식 3) } } Make_Similarity_Matrix_of_User(); Choose_num_of_neighborhood_User(k, user_a); Predict_Missing_Rating_about_User(user_a); //(수식 4) </pre>

첫 번째 단계에서 사용자 a 와 사용자 u 의 유사도 가중치인 $w_{a,u}$ 는 다음과 같이 구할 수 있다.

$$w_{a,u} = \frac{\sum_{i \in I} (R_{a,i} - \bar{R}_a)(R_{u,i} - \bar{R}_u)}{\sqrt{\sum_{i \in I} (R_{a,i} - \bar{R}_a)^2} \sqrt{\sum_{i \in I} (R_{u,i} - \bar{R}_u)^2}} \quad (\text{수식 3})$$

I 가 사용자 a 와 u 모두에 의해 평가된 적이 있는 아이템들의 집합일 때, $R_{u,i}$ 는 사용자 u 의 아이템 i 에 대한 선호도이고, $R_{a,i}$ 는 사용자 a 의 아이템 i 에 대한 선호도이다. 또한, \bar{R}_u 는 사용자 u 의 아이템 집합 I 에 대한 선호도의 평균이며 \bar{R}_a 는 사용자 a 의 아이템 집합 I 에 대한 선호도의 평균이다.

두 번째 단계에서는, 만약 사용자 a 의 아이템 m 에 대한 평가가 빠져 있다면 사용자 a 와 가장 유사도가 높은 k 개의 사용자를 유사도 테이블을 바탕으로 선택한다.

세 번째 단계에서는 사용자 a 의 아이템 m 에 대한 선호도를 사용자 가중치의 평균을 이용해 예측한다.

$$p_{a,m} = \bar{R}_a + \frac{\sum_{k \in K} (R_{k,m} - \bar{R}_a) w_{a,k}}{\sum_{k \in K} w_{a,k}} \quad (\text{수식 4})$$

K 가 사용자 a 와 가장 유사한 k 개의 사용자들의 집합일 때, $p_{a,m}$ 은 사용자 a 의 아이템 m 에 대한 선호도가 된다.

사용자 기반 선호도 예측 기법의 연구사례로, Y. Song가 제안한 군집화 기법을 적용하여 사용자 선호도 예측의 효율성을 높이는 방법이 있다[2].

III. 데이터 신뢰도 기반 가중치를 이용한 하이브리드 선호도 예측 기법

하이브리드 선호도 예측 알고리즘은 아이템 기반 선호도 예측 알고리즘과 사용자 기반 선호도 예측 알고리즘을 하나로 통합한 알고리즘이다. 이는 H. Ji에 의해 처음 제안되었으며 이 연구에서는 이 알고리즘에서 아이템 기반 선호도 예측 결과와 사용자 기반 선호도 예측 결과를 조합하는 과정을 개선한 형태의 알고리즘을 제안하고 있다. 이 알고리즘에서는 사용자 유사도 표와 아이템 유사도 표 모두를 최종 선호도 예측에 이용한다. 또한, 사용자와 사용자간의 관계와 아이템과 아이템간의 관계 모두를 동시에 사용한다. 이와 같은 아이템 기반 선호도 예측 알고리즘과 사용자 기반 선호도 예측 알고리즘의 통합은 예측 정확도의 향상뿐만 아니라, 추천 시스템의 초기 데이터 희소성 문제에 대한 강건성 또한 높일 수 있다. 본 연구에서는 하이브리드 선호도 예측 알고리즘에 데이터 신뢰도에 따른 유동 가중치를 적용하여 데이터 예측 신뢰도를 높인 알고리즘을 제안한다. 표 3은 이 연구에서 제안하는 “데이터 신뢰도 기반 가중치를 이용한 하이브리드 선호도 예측

기법”의 전체 알고리즘에 대한 요약이다.

표 3. 데이터 신뢰도 기반 가중치를 이용한 하이브리드 선호도 예측 알고리즘

Table 3. Hybrid Preference Prediction Algorithm Using Data Reliability based Weighting

데이터 신뢰도 기반 가중치를 이용한 하이브리드 선호도 예측 알고리즘
item_i, item_j: i번째와 j번째의 임의의 아이템 user_i, user_j: i번째와 j번째의 임의의 사용자 item_m: 추천의 대상이 되는 아이템 user_a: 추천을 요청하는 사용자
<pre> for(int i=0; i<num_of_item; i++){ for(int j=0; j<num_of_item; j++){ Calculate_Item_Similarity(item_i, item_j); //(수식 5) } } Make_Similarity_Matrix_of_Item(); for(int i=0; i<num_of_item; i++){ for(int j=0; j<num_of_item; j++){ Calculate_User_Similarity(user_i, user_j); //(수식 6) } } Make_Similarity_Matrix_of_User(); Choose_num_of_neighborhood_Item(k, item_m); Choose_num_of_neighborhood_User(k, user_a); Predict_Missing_Rating_about_Item(item_m); //(수식 8) Predict_Missing_Rating_about_User(user_a); //(수식 9) Merge_Results_of_Rating_Predictions(); //(수식 7) </pre>

첫 번째 단계에서 아이템들 간의 아이템 유사도 가중치는 (수식 5)를 통해서, 사용자들 간의 사용자 유사도 가중치는 (수식 6)을 통해서 계산된다. 아이템 유사도 표와 사용자 유사도 표는 각 두 아이템의 유사도 가중치와 각 두 사용자의 유사도 가중치를 바탕으로 각각 구성된다.

$$w_{i,j} = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad (\text{수식 5})$$

U 가 아이템 i 와 j 를 모두를 평가한 적이 있는 사용자들의 집합일 때, $R_{u,i}$ 는 사용자 u 의 아이템 i 에 대한 선호도이고, $R_{u,j}$ 는 사용자 u 의 아이템 j 에 대한 선호도이다. 또한, \bar{R}_i 는 사용자 집합 U 의 아이템 i 에 대한 선호도의 평균이며 \bar{R}_j 는 사용자 집합 U 의 아이템 j 에 대한 선호도의 평균이다.

$$w_{a,u} = \frac{\sum_{i \in I} (R_{a,i} - \bar{R}_a)(R_{u,i} - \bar{R}_u)}{\sqrt{\sum_{i \in I} (R_{a,i} - \bar{R}_a)^2} \sqrt{\sum_{i \in I} (R_{u,i} - \bar{R}_u)^2}} \quad (\text{수식 6})$$

I 가 사용자 a 와 u 모두에 의해 평가된 적이 있는 아이템들의 집합일 때, $R_{u,i}$ 는 사용자 u 의 아이템 i 에 대한 선호도이고, $R_{a,i}$ 는 사용자 a 의 아이템 i 에 대한 선호도이다. 또한, \bar{R}_u 는 사용자 u 의 아이템 집합 I 에 대한 선호도의 평균이며 \bar{R}_a 는 사용자 a 의 아이템 집합 I 에 대한 선호도의 평균이다.

두 번째 단계에서는 사용자 a 의 아이템 m 에 대한 평가가 빠졌을 경우 사용자 a 를 대상 사용자로 아이템 m 을 대상 아이템으로 한다. 이 과정에서는 아이템 m 과 가장 유사도가 높은 k 개의 아이템을 유사도 테이블을 바탕으로 선택한다. 또한 동시에 사용자 a 과 가장 유사도가 높은 k 개의 사용자를 유사도 테이블을 바탕으로 선택한다.

세 번째 단계에서는 사용자 a 의 아이템 m 에 대한 평가를 k 개의 사용자 가중치의 평균과 k 개의 아이템 가중치의 평균의 합으로 예측하며, 그 과정은 (수식 7)과 같다.

$$p_{a,m}(\text{hybrid}) = \alpha \times p_{a,m}(u) + (1 - \alpha) \times p_{a,m}(i) \quad (\text{수식 7})$$

이때, $p_{a,m}(u)$ 는 사용자 가중치의 평균을 의미하며, $p_{a,m}(i)$ 은 아이템 가중치의 평균을 의미하며, 각각 (수식 8)과 (수식 9)을 통해 계산된다.

$$p_{a,m}(u) = \bar{R}_a + \frac{\sum_{k \in K} (R_{k,m} - \bar{R}_a)w_{a,k}}{\sum_{k \in K} w_{a,k}} \quad (\text{수식 8})$$

K 가 사용자 a 와 가장 유사한 k 개의 사용자들의 집합일 때, $p_{a,m}$ 은 사용자 a 의 아이템 m 에 대한 선호도가 된다.

$$p_{a,m}(i) = \frac{\sum_{k \in K} R_{a,k}w_{m,k}}{\sum_{k \in K} w_{m,k}} \quad (\text{수식 9})$$

K 가 아이템 m 과 가장 유사한 k 개의 아이템들의 집합일 때, $p_{a,m}$ 은 사용자 a 의 아이템 m 에 대한 선호도가 된다.

α 는 예측 결과 값에 대한, 아이템 기반 선호도 예측 결과와 사용자 기반 선호도 예측 결과의 비율을 의미한다. 본 연

구에서는 Hao Ji[4]의 연구에서 고정된 특정 값으로 정의된 α 를 모델의 상황에 따라 능동적으로 변화하는 값으로 정의한다. α 를 구하는 방법은 (수식 10)과 같다.

$$\alpha = \frac{\sigma(K(i))}{\sigma(K(u)) + \sigma(K(i))} \quad (\text{수식 10})$$

이때, $K(u)$ 는 사용자 a 에 대한 유사 사용자 집합을 $K(i)$ 는 아이템 m 에 대한 유사 아이템 집합을 $\sigma(K(u))$ 는 유사 사용자 집합의 표준 편차를 $\sigma(K(i))$ 는 유사 아이템 집합의 표준편차를 나타낸다. 이러한 방법은 각 유사 사용자 집합과 유사 아이템 집합의 선호도 예측의 근거로써의 신뢰도를 표준 편차의 형태로 추정하여 사용자 기반 선호도 예측과 아이템 기반 선호도 예측의 결과 값을 조합하는데 반영함으로써 예측의 정확도를 높일 수 있다.

IV. 실험 및 검증

본 연구에 대한 검증은 추천 시스템의 성능 평가를 위해 가장 많이 사용되는 평균절대오차(MAE)를 이용한다[1,5]. 이 측정법은 추천 시스템이 예측한 순위와 실제 순위의 차이를 보여준다. 평균절대오차(MAE)를 구하는 방법은 (수식 11)과 같다.

$$MAE = \frac{\sum_{i=1}^N |p_i - r_i|}{N} \quad (\text{수식 11})$$

여기서 N 은 예측한 아이템의 개수이고, p_i 는 번째 아이템의 예측한 순위, r_i 는 번째 아이템의 실제 순위이다.

실험은 다음과 같은 환경에서 진행되었다. 서버는 Windows 7기반의 Apache Tomcat 7.0을 사용하였으며, 데이터베이스는 MySQL5.5를 사용하였다. 서버 측 개발환경은 VisualStudio 2010를 사용하였으며, 사용된 서버 측 언어는 VC++이고, 클라이언트는 안드로이드 응용프로그램으로 구현되었으며, 클라이언트 측 통합개발환경은 Eclipse Indigo, 개발언어는 Android SDK를 이용한 JAVA이다.

실험 데이터는 경기도 인근의 300여 개의 문화재, 관광지, 레저 콘텐츠를 이용하였다. 아이템 제공 및 추천은 아이템은 사용자가 지정한 거리 내의 인근 아이템을 사용자의 프로필과, 상황정보에 따라 추천하여 아이템의 이용이 끝나면 0점이

서 10점까지의 점수를 입력하여, 이용한 아이템에 대해 평가하는 방식으로 이루어진다.

실험을 위해 구현된 추천 시스템은 모바일 환경에서 사용자들에게 증강현실 기반의 문화재, 레저, 관광지 콘텐츠를 증강현실로 제공하기 위해 구축되었다. 구현된 추천 시스템의 모바일 단말 상에서의 실행 화면은 그림 1과 같다. 이 두 화면은 각각 증강현실 형태와 지도 형태의 인터페이스를 통해 사용자에게 아이템을 제공하는 화면이다.



그림 1. 구현된 추천 시스템의 모바일 단말 상의 실행 화면
Fig. 1. Screen Shot of Implemented Recommendation System in Mobile Device

실험은 추천 서비스의 신뢰도 개선을 검증하기 위하여, 아이템 기반 협업 필터링(ICF, Item Based Collaborative Filtering), 사용자 기반 협업 필터링(UCF, User Based Collaborative Filtering), Hao Ji가 제안한 선호도 예측 기법[4]인 HF(Hybrid collaborative Filtering)를 적용한 협업 필터링 기법을 본 연구에서 제안한 기법인 HFMC(Hybrid collaborative Filtering applying Model Condition)와 같은 환경에서 구현하여 평균절대오차(MAE)를 통해 비교하였다.

실험 절차는 다음과 같다. 1000명 이상의 사용자가 이용하는 추천 시스템에서 사용자 선호도 행렬의 희박도에 따른 추천 서비스 신뢰도의 변화를 측정하는 방식으로 진행되었다. 실험 대상은 단국대학교 학생 및 교직원, 인근주민 중 무작위적으로 추출된 50명의 표본 사용자 집단을 대상으로 하였다. 연령 분포는 20대 15명, 30대 14명, 40대 10명, 50대 9명으로 균등하게 구성하였다.

실험은 30개의 아이템을 추천하고 모든 아이템에 대한 선호도 정보를 입력 받아, 사용자가 평가한 순위와 추천 시스템

에 의해 예측된 순위를 비교하는 방식으로 진행되었다. 데이터 수집은 선호도 행렬 희박도가 98%인 상황에서부터 80%인 상황까지 2% 간격으로 10가지 상황에서 수집하였다. 그림 2는 각 기법을 적용한 추천 시스템에서 사용자 피드백 수의 변화에 따른 평균절대오차(MAE)의 변화를 나타낸 그래프이다.

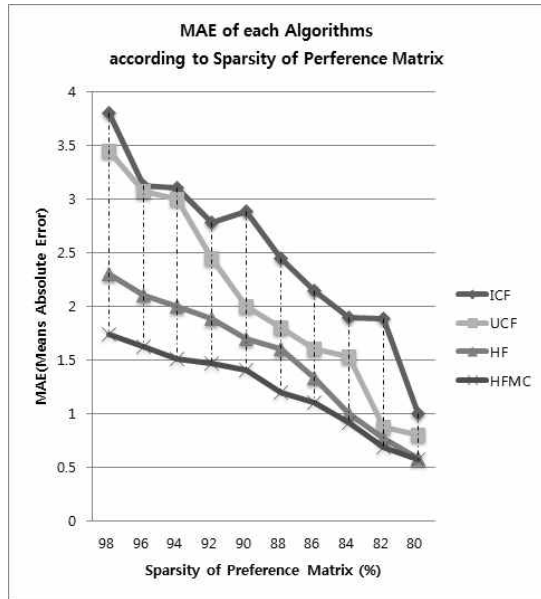


그림 2. 선호도 행렬 희박도 변화에 따른 알고리즘 별 MAE
Fig. 2. MAE of Each Algorithms according to Sparsity of Preference Matrix

그림 2는 보면 모든 추천 시스템들이 사용자 피드백의 수가 늘어날수록 점차 추천 성능이 높아지는 것을 볼 수 있다. 제한한 선호도 예측 기법을 적용한 추천 시스템의 경우, 선호도 정보가 충분한 선호도 행렬의 희박도가 84% 이하인 상황에서는 다른 선호도 예측 기법들을 이용한 추천 시스템들과 비슷한 추천 신뢰도를 보인다. 하지만 선호도 행렬의 희박도가 84% 이상인 상황에서는 사용자 피드백이 적어 HF 기법을 이용한 추천 시스템에 비해 평균절대오차(MAE) 측면에서 평균 21.7% 개선된 성능을 보인다. 이는 본 논문에서 제안하는 기법이 추천 시스템의 전반적인 성능을 끌어올리지는 못하지만, 기존 협업 필터링 기법의 행렬 희박성 문제를 개선하였다는 것을 증명한다.

V. 결론

본 연구에서는 모델 상황을 반영한 선호도 예측 기법을 제안하고 이를 사용한 추천 시스템을 구현하였다. 또한 실제 문화재, 관광지, 레저 콘텐츠 추천 서비스에 적용하여 그 성능을 평가함으로써 제안하는 기법의 유효성을 검증하였다.

제안하는 모델 상황을 반영한 선호도 예측 기법은 협업 필터링 추천 기법에서 나타나는 사용자 선호도 행렬의 희박성 문제를 해결하기 위한, 기존 선호도 예측 기법보다 높은 정확도를 보일 수 있는 선호도 예측 기법이다. 이는 아이템 기반 선호도 예측 기법과 사용자 기반 선호도 예측 기법을 병용하고 이 두 기법의 예측치를 각 유사 아이템, 사용자 집합의 신뢰도를 표준편차를 이용해 추정된 가중치 값을 통해 병합함으로써 이뤄진다.

본 연구는 기존 아이템 기반 선호도 예측 기법, 사용자 기반 선호도 예측 기법, Hao Ji의 기법[2]과 본 연구에서 제안한 선호도 예측 기법을 각각 협업 필터링 추천 시스템에 적용하여 비교 평가하였다. 평균절대오차(MAE)를 이용한 추천 정확도 실험은 기존 Hao Ji의 기법에 비해 선호도 행렬 희박도가 84% 이상인 상황에서 평균 21.7% 개선된 평균절대오차(MAE)를 보여, 모델 상황을 반영한 선호도 예측 기법이 기존 선호도 예측 기법들에 비해 선호도 행렬 희박성 문제를 효과적으로 개선할 수 있음을 증명하고 있다.

본 논문은 확장성 측면에서 개선할 필요가 있다. 행렬 희박성 문제를 개선하였지만, 초기 평가, 초기 이용자 문제에 대한 해결책을 제시하지 못하고 있으며, 확장성 문제를 더욱 악화시키고 있다. 향후 연구는 위와 같은 문제를 해결하기 위하여, 사용자와 아이템에 대한 군집화를 활용하여 초기 평가, 초기 이용자 문제와 확장성 문제에 대한 해결책을 제시하고, 사용자 상황정보를 활용하여 추천 서비스의 만족도를 높이는 방향으로 나아갈 것이다.

참고문헌

- [1] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, Vol.22, No.1, 2004, pp. 5-53.
- [2] Y. Song, C. L. Giles, "Efficient user preference

- predictions using collaborative filtering," *Pattern Recognition*, 2008. ICPR 2008. 19th International Conference, 2008, pp. 1-4.
- [3] J. Reed, C. H. Lee, "Preference Music Ratings Prediction Using Tokenization and Minimum Classification Error Training," *Audio, Speech, and Language Processing*, IEEE Transactions, Vol. 19, No. 8, 2011, pp. 2294-2303.
- [4] H. Ji, J. Li, C. Ren, and M. He, "Hybrid collaborative filtering model for improved recommendation," *Service Operations and Logistics, and Informatics (SOLI)*, 2013 IEEE International Conference, 2013, pp. 142-145.
- [5] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: a constant time collaborative filtering algorithm," *Information Retrieval*, Vol.4, No.2, 2001, pp. 133-151.
- [6] M. O. Connor, J. Herlocker, "Clustering Items for Collaborative Filtering," *Proc. of the ACM SIGIR Workshop on Recommender Systems*, Berkeley, CA, 1999.
- [7] G. Groh, C. Ehming, "Recommendations in Taste Related Domains: Collaborative filtering vs. Social filtering," *Proc. of GROUP'07*, 2007, pp. 127-136.
- [8] J. L. Herlocker, J. A. Konstan, and J. T. Riedl, "An Empirical Analysis of Design Choices in Neighborhood-based Collaborative Filtering Systems," *Information Retrieval*, Vol. 5, 2002, pp. 287-310.
- [9] Z. Huang, H. Chen, and D. Zeng, "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," *ACM Trans. Information Systems*, Vol.22, No.1, 2004, pp. 116-142.
- [10] J. Konstan, D. B. Miller, and D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying collaborative filtering to Usenet news," *Communication of ACM*, Vol.40, No.3, 1997, pp. 77-87.
- [11] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithm," *Proc. of the 10th international conference on World Wide Web*, 2001, pp. 285-295.
- [12] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-Boosted Collaborative Filtering for Improved Recommendations," *Proc. of the Eighteenth National Conference on Artificial Intelligence*, Edmonton, Canada, 2002, pp. 187-192.
- [13] M. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," *Artificial Intelligence Review*, 1999, pp. 398-408.
- [14] B. Jin, Y. Cho, and K. Ryu, "Personalized e-commerce recommendation system using RFM method and association rules," *Journal of the Korea Society of Computer and Information*, Vol.15, No.12, 2010, pp. 227-235.
- [15] D. Park, "Improved Movie Recommendation System based-on Personal Propensity and Collaborative Filtering," *KIPS transactions on computer and communication systems*, Vol.2, No.11, 2013, pp. 475-482.
- [16] S. B. Park, Y. T. Baek "A Study of Story Visualization Based on Variation of Characters Relationship by Time," *Journal of the Korea Society of Computer and Information*, Vol.18, No.3, 2013, pp. 119-126.
- [17] W. J. Lee, K. W. Kim, K. D. Boo, and J. J. Woo, "A Study on the Adoption of NAC for Guaranteeing Reliability of u-Campus Network," *Journal of Korean Institute of Information Technology*, Vol.7, No.4, 2009, pp. 252-258.
- [18] E. S. You, S. B. Park, "Story-based Information Retrieval," *Journal of Intelligence and Information Systems*, Vol.19, No.4, 2013, pp. 81-96.

저 자 소 개



이 오 준

현 재 : 단국대학교

소프트웨어학과 학부 재학중

관심분야: 추천시스템, 기계학습,

정보검색, 적응형시스템

Email : concerto_grs@naver.com



백 영 태

1989: 인하대학교

전자계산학과 이학사

1993: 인하대학교

전자계산공학과 공학석사

2002: 인하대학교

전자계산공학과 공학박사

1993-1998: 대상정보기술(주)

정보통신연구소

선임연구원

1998-현 재: 김포대학교

멀티미디어과 부교수

관심분야: 멀티미디어 정보검색,

웹교육시스템, 모바일시스템

Email : hanna@kimpo.ac.kr