

협력필터링 시스템을 위한 평가 등급 범위 기반의 예측방법

이수정[†]

요 약

인터넷 상에서 사용자 흥미에 부합하는 항목을 예측하여 추천해 주는 추천 시스템은 e-commerce가 발달함에 따라 다양한 분야에서 적용되어 왔다. 추천 시스템의 주요 방법인 협력 필터링은 사용자가 선호했던 항목들과 유사한 항목을 추천하거나 또는 유사한 기호의 다른 사용자가 선호했던 항목을 추천하는 것이다. 따라서 유사도의 정확한 측정과 추천한 항목의 실제 평가등급 예측은 협력 필터링의 성능을 결정하는 두가지 중요한 문제이다. 본 연구에서는 후자의 문제를 다룬다. 기존 연구에서는 평가 등급의 평균값을 기반으로 하여 실제 평가등급을 예측하였으나, 본 연구에서는 평가 등급 범위 기반의 방법을 제시하고 실험을 통해 성능을 조사하였다. 실험 결과 기존 방법에 비해 제안 방법은 평균 절대 오차에 있어서 성능이 크게 향상됨을 입증하였다.

주제어 : 추천 시스템, 웹 개인화, 협력 필터링, 유사도

A Rating Range-based Prediction Method for Collaborative Filtering Systems

Soojung Lee[†]

ABSTRACT

Recommender systems, which predict and recommend items that may possibly draw users' interests, have been applied in various fields as e-commerce systems are widespread. Collaborative filtering, one of the major methodologies of recommender systems, recommends either items similar to those preferred by the user, or items preferred by the other similar user. Therefore, two problems determine its performance; one is correct estimation of similarity and the other is predicting the real rating of the recommended item. This study addresses the latter problem. Previous studies predict the real rating based on the mean of the ratings, but this study proposes a prediction based on the range of the ratings and investigates its performance through experiments. As a result, it is demonstrated that the proposed method improves the mean absolute error significantly, compared to the previous method.

Keywords : Recommender System, Web Personalization, Collaborative Filtering, Similarity

[†] 정 회 원: 경인교육대학교 컴퓨터교육과 교수
논문접수: 2011년 06월 23일, 심사완료: 2011년 07월 21일

1. 서론

웹 문서 수가 급증함에 따라 인터넷에서 원하는 정보를 빠른 시간 내에 검색하기가 점차 어려워짐에 따라 거의 90%의 검색 결과가 사용자가 원하지 않는 불필요한 것으로 밝혀졌다[1]. 이러한 문제를 경감시키기 위해 사용자 개인의 서로 다른 기호에 부합하는 정보를 손쉽게 찾을 수 있도록 하는 웹 개인화 시스템이 각광받고 있다. 웹 개인화의 주를 이루는 예는 추천 시스템으로서 주로 고객이 원하는 상품, 영화, 뉴스, 서적, 여행 정보 등 다양한 주제로 개발되었다[2][3].

추천 시스템의 주요한 두가지 방법은 내용기반 필터링(content-based filtering)과 협력 필터링(collaborative filtering)이다[4]. 내용기반 필터링은 내용 분석을 토대로, 개인적인 사용자 흥미도 프로필을 구축하여 그에 부합하는 정보를 추천한다. 프로필의 구축은 사용자가 직접 입력하거나 또는 사용자의 행위로부터 간접적으로 학습한다. 예를 들어, WebMate[5]는 사용자가 흥미를 보이는 문서들로부터 간접적이고 자동적으로 사용자의 선호 영역을 학습하고, [6]에서 제안한 시스템은 사용자의 검색 이력으로부터 선호 범주를 학습한다.

이에 반하여 협력 필터링은 여러 다른 사용자들의 의견에 따라 웹 사이트를 추천하는 방법으로서 많은 e-commerce 시스템에서 성공적으로 적용하고 있다. 초창기의 Tapestry 시스템[7]은 유사한 사용자들을 수작업으로 구분하였으며, 이후 개발된 GroupLens[8], Video Recommender[9]는 사용자의 흥미도를 자동으로 예측하기 위한 시스템들이다. 이밖에 서적을 추천해 주는 협력 시스템인 Amazon.com, WWW 상에서 관련 있는 정보를 찾아 주는 PHOAKS system[10], 농담을 찾아 주는 Jester system[11] 등이 있다.

협력 필터링을 이용한 추천 시스템은 크게 메모리 기반(memory-based) 또는 휴리스틱 기반(heuristic-based)과 모델 기반(model-based)의 두가지로 분류된다[2]. 메모리 기반 시스템은 두 명의 사용자가 유사한 기호를 갖고 있다면, 다른 새로운 항목들에 대해서도 유사한 반응을 보일 것이라는 가정에 기반한다. 따라서 항목들에 대해

사용자들이 과거에 부여한 등급 기록을 이용한다. 이에 반해 모델 기반 알고리즘은[12][13] 과거 등급들을 이용하여 모델을 학습하고 이를 토대로 새로운 항목에 대한 등급을 예측한다. 모델의 학습은 주로 확률, 통계나 기계 학습 기법을 활용하지만 좀더 복잡한 기법으로서 Bayesian, 선형 회귀분석, 최대 엔트로피 모델, 마코프 결정 프로세스(Markov decision processes) 등을 활용하였다.

본 논문에서는 메모리 기반 협력 필터링을 통한 추천 시스템 상에서 새로운 항목에 대한 등급 예측을 위한 방법을 연구한다. 기존에 주로 사용했던 등급 평균을 이용한 방법의 문제점을 분석하고, 이를 보완하기 위해 각 개인이 부여한 등급의 범위를 기반으로 예측하는 방법을 제안한다. 논문의 구성은 다음과 같다. 2절에서는 관련 지식을 기술하고, 3절에서 본 연구에서 제안하는 예측 방법을 설명하며 4절에서 실험을 통한 성능을 입증하고 5절에서 논문의 결론을 맺는다.

2. 기존 연구

2.1 메모리 기반 협력 필터링

메모리 기반의 협력 필터링을 통한 추천 시스템은 사용자에게 새로운 항목들에 대해 그가 부여할 등급을 예측하여, 가장 높은 예측등급의 항목을 추천한다. 따라서 가능한 한 정확히 예측하는 것이 관건이다. 이를 위하여 다음과 같은 절차를 따른다. 우선 과거에 사용자들이 부여하였던 모든 등급 이력을 보관하고, 공통된 항목에 대해 부여하였던 등급 리스트를 토대로 사용자간 또는 항목간 유사도를 계산한다. 메모리 기반 협력 필터링 시스템은 사용자 기반과 항목 기반의 두가지로 분류하는데[14], 사용자 기반(user-based)의 시스템인 경우, N명의 최인접 이웃이 부여한 등급의 최고값에 해당하는 항목을 추천하고, 항목 기반(item-based)의 시스템인 경우는 사용자가 선호하였던 항목들과 가장 유사한 항목을 추천한다.

2.2 유사도

유사도 측정은 최인접 이웃을 알아내는 중요한

역할을 하므로 협력 필터링의 성능을 좌우한다. 현재까지 가장 널리 사용되는 측정방법으로 Pearson 상관계수와 cosine similarity가 있다[15]. 사용자 u와 v가 모두 등급을 매긴 항목들의 집합을 I라고 할 때, 두 사용자 간의 Pearson 상관계수는

$$sim(u, v) = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \dots \text{식 <1>}$$

으로 산출한다. 이 때 $r_{u,i}$ 는 사용자 u가 부여한 항목 i에 대한 등급이고, $r_{v,i}$ 는 사용자 v가 부여한 항목 i에 대한 등급이다. \bar{r}_u 와 \bar{r}_v 는 각각 사용자 u와 v가 부여한 평균 등급이다.

Cosine similarity는 각 사용자를 II 차원의 벡터로 간주하여 두 벡터간 각도의 cosine 값으로써 유사도를 측정한다. 즉,

$$sim(u, v) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \times \|\vec{v}\|} = \frac{\sum_{i \in I} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I} r_{u,i}^2} \sqrt{\sum_{i \in I} r_{v,i}^2}} \dots \text{식 <2>}$$

기타 다른 유사도 측정 방법은 [16]에 정리되어 있으므로 이 논문에선 생략한다.

2.3 등급 예측

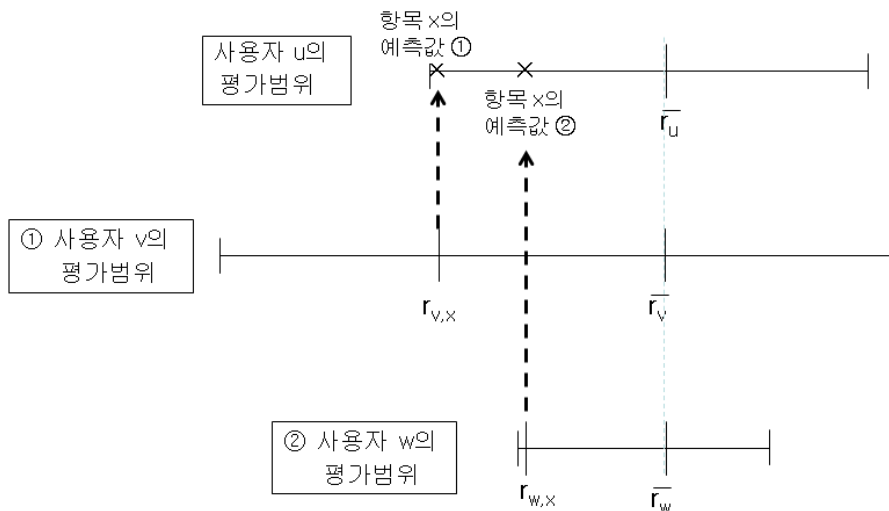
사용자 기반의 협력 필터링은 가장 유사한 사용자들로부터 그들이 부여한 등급 정보를 취합하여, 높은 등급을 얻은 항목을 추천하는 시스템이다. 따라서 추천한 항목의 등급이 사용자가 부여할 실제 등급에 얼마나 부합하는지의 예측 정확도가 매우 중요하다. 사용자 u와 인접한 사용자들의 집합을 N_u 이라 할 때, 사용자 u가 등급을 부여하지 않은 항목 x에 대한 예측 등급 $r_{u,x}$ 는 대개 다음과 같이 산출한다[4][15].

$$r_{u,x} = \bar{r}_u + \frac{\sum_{v \in N_u} sim(u, v) \times (r_{v,x} - \bar{r}_v)}{\sum_{v \in N_u} |sim(u, v)|} \dots \text{식 <3>}$$

3. 제안 예측 방법

3.1 기존 방법의 문제점

<그림 1>은 사용자 u가 아직 평가하지 않은 항목 x에 대해, 다른 인접 사용자 v와 w의 평가 등급을 이용하여 기존 방법인 식 <3>에 따라 사용자 u의 평가등급을 예측하는 두 가지 예를 보여준다. 각 사용자별로 가로선은 해당 사용자가 모든 항목에 대해 부여한 평가등급의 범위를 나



<그림 1> 기존 예측 방법의 문제점 예시

타낸다. 사용자 i 가 부여한 평가등급의 평균은 \bar{r}_i 로 표기하였고, 항목 x 에 대해 사용자 v 와 w 가 각각 부여한 평가등급은 $r_{v,x}$ 와 $r_{w,x}$ 로 표기하였다.

그림에서 ①의 경우에 인접 사용자 v 의 평가등급에 따르면 사용자 u 의 x 에 대한 평가등급의 예측값은 사용자 u 의 평가범위 선상에 표시된 예측값 ① 부근이 될 것이다. 물론 이때 인접 사용자는 v 한명이고 u 와 v 의 유사도는 고려하지 않음을 가정했다. 그런데 이 예측값은 사용자 u 에 있어서는 최저 평가에 가깝지만, 사용자 v 에 있어서는 그렇지 않고 평균 \bar{r}_v 보다 다소 낮은 등급임을 알 수 있다. 즉, 항목 x 에 대한 사용자 u 와 v 의 평가의도는 매우 상이함을 보여준다. 이와는 반대로 ②의 경우에 $r_{w,x}$ 는 사용자 w 에 있어서 거의 최저등급인 반면 사용자 u 에 있어서 예측값 ②는 그렇지 않음을 볼 수 있다.

인접 사용자의 평가가 제대로 반영되었다면, 예측값 ①은 사용자 u 의 평가범위 선상에서 보다 우측으로, 예측값 ②는 거의 최저등급에 위치해야 한다. 왜냐하면 사용자 w 는 항목 x 에게 거의 최저등급을, 사용자 v 는 최저가 아닌 평균 이하등급을 부여했기 때문이다.

3.2 제안 방법

앞절에서 언급한 문제점을 해결하기 위한 한가지 방법으로 본 연구에서는 평가범위를 이용한 예측방법을 제시하려 한다. 즉, 사용자가 부여한 모든 평가등급들 중에서 항목 x 의 평가는 어느 정도에 위치하는지를 고려하고, 이러한 상대적 위치 정보를 인접 사용자의 평가등급 산출에 반영한다. 따라서, 예측 평가 등급은 다음과 같다.

$$r_{u,x} = \min_u + W \sum_{v \in N_u} \frac{range_u}{range_v} (r_{v,x} - \min_v) .$$

.....식 <4>

이 식에서 \min_u 는 \min_v 는 각각 사용자 u 와 v 의 최저평가등급이며, W 는 가중치, N_u 는 사용자 u 와 인접한 사용자들의 집합, $range_u$ 와 $range_v$ 는 각각 해당 사용자가 부여한 모든 평가등급의 범위이다.

가중치 W 는 다음과 같은 경우를 고려하여 결

정한다. <그림 1>에서 만약 사용자 w 가 평가등급을 높게 주는 경향이 있는데, 그가 부여한 평가범위가 매우 작다면 $r_{w,x}$ 는 사용자 w 에게는 최저등급에 가깝지만, 타사용자들에게는 매우 높은 등급일 수 있다. 그러나, 인접 사용자 w 의 등급을 반영하기 위해, 식 <4>에서 가중치 W 를 제외하고 적용했을 때, 예측값 $r_{u,x}$ 는 사용자 u 에게 있어서도 최저등급에 가까울 것이다. 만약 사용자 u 의 평가범위가 사용자 w 의 평가범위보다 상대적으로 매우 크다면, 이 예측값은 사용자 w 의 의도를 제대로 반영하지 않은 것이다. 따라서 인접 사용자의 평가범위의 상대적 크기를 가중치 산출식의 요소로 하여 다음과 같이 결정한다.

$$W = \left(\frac{range_v}{range_u} \right) / \left(\sum_{v \in N_u} \frac{range_v}{range_u} \right)$$

4. 실험 연구

4.1 실험 배경

제안한 예측 방법의 성능을 평가하기 위하여, Book-Crossing 데이터를 사용하였다[17]. 이 데이터는 271,379권의 서적에 대하여 278,858명의 사용자들이 평가한 1-10 사이의 등급을 1,149,790개 포함하고 있다. 이들 중 많은 사용자 또는 서적의 평가수가 매우 적어 성능 측정의 정확성이 떨어지므로, 원 데이터 행렬의 희소수준을 떨어뜨림으로써 보다 정확한 평가가 되도록 하기 위하여 사전 처리하였다. 결과적으로 [15]에서처럼 희소수준이 다른 두 종류의 데이터 집합을 산출하였는데, 첫째 집합에 속한 각 서적에 대해서는 적어도 11명의 사용자가 등급을 매겼으며, 또한 이 집합의 각 사용자는 적어도 6개의 서적에 대해 평가하였다. 이 집합을 Book-Crossing1이라고 표기하기로 한다. 둘째 데이터 집합의 각 서적에 대해서는 적어도 21명의 사용자가 평가를 내렸으며, 이 집합의 각 사용자는 첫째 집합과 마찬가지로 적어도 6개의 서적에 대해 평가하였다. 이 집합을 Book-Crossing2라고 명명하였다. 각 데이터 집합의 크기에 대해 <표 1>에 상세 기술하였다. 희소성 수준이란 행렬 내 데이터가 없는 요소, 즉, 평

가가 매겨지지 않은 요소의 비율을 의미하며, (값이 0인 요소 개수)/(행렬의 크기)로 산출하였다.

<표 1> 실험 데이터 집합

	Book-Crossing1	Book-Crossing2
평가개수	48456	17841
행렬크기 (사용자수×서적수)	3072 × 2097	1498 × 480
희소성수준	0.992478	0.975188

사용자가 부여한 평가등급에 대한 예측의 정확도는 여러 가지 요인에 의해 결정되는데, 본 논문에서는 기존에 가장 많이 활용되었던 두 종류의 유사도 계산 방식, 즉, Pearson correlation과 cosine similarity의 영향을 실험하였다. 이 밖에 예측 성능에 대한 인접 사용자 수(Number of nearest neighbors)의 영향을 조사하기 위하여, 인접 사용자 수를 5부터 35까지 변화시켜 실험하였다. 앞 절에서 설명한 기존의 예측 방법을 'Prev', 본 논문에서 제안한 방법을 'Prop'라고 표기하였다. 성능 평가 척도로서 관련 연구에서 주로 사용하는 MAE (Mean Absolute Error)[18]를 도입하였고 이 척도는 다음과 같이 정의된다.

$$MAE = \frac{\sum_x |r_{u,x} - p_{u,x}|}{N}$$

$r_{u,x}$: 서적 x에 대해 사용자 u가 부여한 등급
 $p_{u,x}$: 서적 x에 대한 사용자 u의 등급 예측값

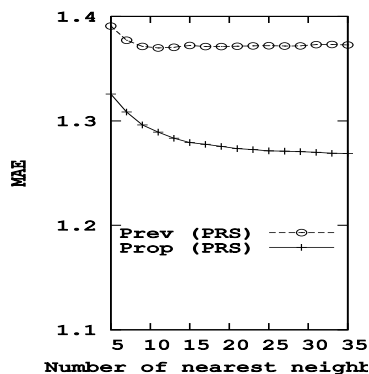
N : 성능 평가 대상 서적의 총개수

실험 결과의 신뢰도를 높이기 위하여 각 실험의 결과 MAE는 10회 크로스 확인(10-fold cross validation)의 평균값으로 산출하였는데, 각 회마다 서로 다른 훈련 데이터 집합과 시험 데이터 집합을 80:20으로 구성하였다. 모든 실험은 1.96GM RAM과 3.16GHz Intel Core 2 Duo CPU의 PC 상에서 C 프로그램을 작성하여 진행하였다.

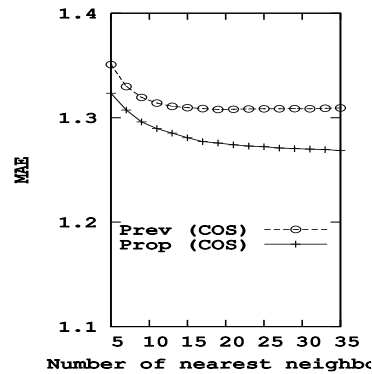
4.2 실험 결과

<그림 2>는 Pearson 상관도와 cosine similarity를 각각 사용하여 최인접 사용자들을 결정한 후, 이들이 부여한 등급을 기초로 예측하였을 때 그 예측등급의 정확도를 비교한 그래프이다. 실험은 Book-Crossing1 데이터 집합에 대해 진행하였고, 인접 사용자수를 5부터 35까지 변화시켜 성능에 대한 영향 정도를 알아보았다.

그림에서 보듯이 두 방법의 성능은 Pearson 상관도와 cosine 상관도를 사용한 각 결과에서 대체적으로 유사한 행태를 보인다. 단, 차이점은 기존 방법인 Prev에서는 인접 사용자수가 약 9~13인 경우부터 더 이상의 성능 진전이 없으나, Prop에서는 인접 사용자수의 증가에 따라 미세하게나마 꾸준히 성능이 향상되었다. Prev의 성능 결과에서 알 수 있는 것은 인접 사용자 수를 더 많이



(a)

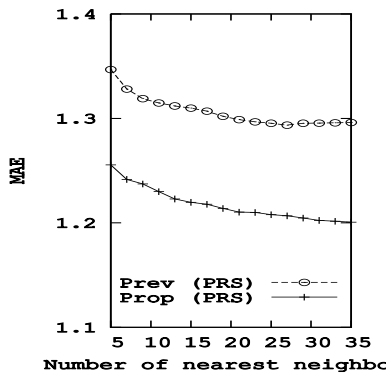


(b)

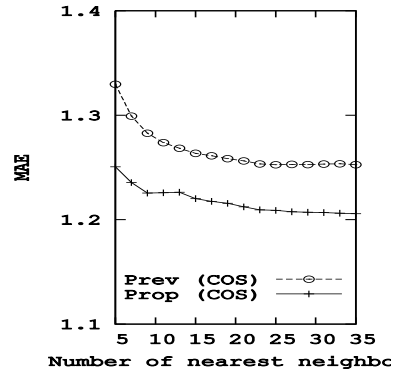
<그림 2> Book-Crossing1 데이터 집합에 대한 각 예측 방법의 MAE 성능.
 (a) Pearson 상관도를 사용한 경우. (b) Cosine similarity를 사용한 경우

<표 2> 인접사용자 순위 그룹별 평균 상관도

인접사용자의 인접순위		1~5	6~10	11~15	16~20	21~25	26~30	31~35
Book-Crossing1	PRS	0.929	0.928	0.922	0.910	0.893	0.873	0.850
	COS	0.995	0.995	0.992	0.989	0.985	0.979	0.973
Book-Crossing2	PRS	0.912	0.912	0.911	0.908	0.902	0.890	0.875
	COS	0.992	0.992	0.992	0.992	0.989	0.985	0.983



(a)



(b)

<그림 3> Book-Crossing2 데이터 집합에 대한 각 예측 방법의 MAE 성능. (a) Pearson 상관도를 사용한 경우. (b) cosine similarity를 사용한 경우

포함할수록 낮은 유사도의 이웃이 더욱 포함되므로 성능 개선에 크게 효과가 없다는 것이다. <표 2>에서 제시한 바와 같이 인접 사용자수에 따라 평균 상관도는 예상한 대로 다소 감소하는 추세를 보였다.

반면에 제안한 방법의 성능은 인접 사용자수가 증가함에 따라 점차 조금씩 좋아지는데, 그 이유는 제안 방법이 유사도를 이용하여 예측하지 않고 각 인접 사용자가 부여한 등급 범위를 활용한 것에 기인한다고 볼 수 있다. 구체적으로 인접 사용자수가 약 20 이상인 경우 기존 방법보다 0.1 정도 낮은 MAE 결과를 나타내었다.

Cosine similarity를 사용한 같은 방식의 실험을 Book-Crossing1 데이터 집합에 실시하여 <그림 2>(b)의 결과를 얻었다. [16]에서 언급하였듯이 Pearson 상관도보다 더 우수한 성능을 나타내었고, 이는 Prev에만 해당됨을 알 수 있다. 그 이유는 제안한 Prop 예측 방법이 앞절의 식 <4>에서 제시한대로 유사도값의 영향을 받지 않기 때문이다. 또한 <표 2>에서 인접순위가 떨어짐에 따라 cosine 상관도의 저하율은 Pearson 상관도에 비해

상대적으로 낮은데도 불구하고, Prev의 경우에 <그림 2>(a)와 (b)가 비슷한 양상을 보임을 주목할 수 있다. 즉, 10명 내외의 인접 사용자수가 성능면에서 최적이며 사용자가 추가되더라도 그들의 평가점은 성능 향상에 크게 영향을 주지 않는 것으로 드러났다. 또한 cosine 상관도를 사용한 경우 초기에 MAE의 향상폭이 상대적으로 더욱 크게 나타났는데, 이는 <표 2>에서 보듯이 cosine 상관도가 Pearson에 비해 더 크기 때문이다. Prev의 경우에 Pearson 상관도를 사용했을 때 보다 cosine similarity를 적용했을 때 대략 0.06 정도의 MAE 향상을 가져왔다.

<그림 3>은 Book-Crossing2 데이터 집합에 대한 실험 결과이다. Book-Crossing1의 결과와 거의 비슷한 양상을 보이나, <그림 3>(a)에서 성능은 두가지 방법 모두 약 0.6~0.7 정도 향상되었음을 알 수 있다. 이는 Book-Crossing2의 경우 희소성 수준이 더욱 나아진 것에 기인한다. 즉, 등급을 부여받은 서적의 수가 증가하게 되므로, 같은 등급을 부여한 인접 사용자의 수도 증가하므로 보다 정확한 등급 예측이 가능해지기 때문이

다. 또한 이러한 같은 등급의 인접 사용자수는 Book-Crossing1의 경우보다 상대적으로 약간 낮은 Pearson 상관도(<표 2> 참조)를 충분히 보상하고도 남는다는 것을 알 수 있다. Prop 방법은 Book-Crossing1의 경우처럼 Prev와 거의 비슷한 정도의 차이를 보여 월등한 성능 향상을 가져왔다.

<그림 3>(b)에서 cosine similarity를 이용하여 Book-Crossing2 데이터 집합에 대해 실험한 결과, Book-Crossing1에 대한 실험 결과에 비해 두 가지 방법 모두 대략 0.4~0.6의 성능 향상을 가져왔으며, 특히 Prop의 경우에 더욱 향상이 컸다. 또한 Prop은 Prev 보다 최대 0.08 향상된 MAE 결과를 나타내었다. 결과적으로 종합해 보면, 제안한 예측 방법은 유사도 측정 방법에 구애받지 않고 Book-Crossing2 데이터 집합에서 가장 좋은 성능을 가져왔으며, 인접 사용자수가 증가함에 따라 성능이 다소 향상되는 경향을 보이나, 증가 숫자가 크면 향상 정도는 미미해짐을 알 수 있었다.

5. 결 론

협력 필터링을 통한 추천 시스템은 광범위한 자료들 중에서 사용자에게 필요할 만한 자료들만을 골라 제시하므로 서적, 뉴스, 영화 등 다양한 분야에서 매우 유용하게 실제로 활용되어 정보화 시대에서 중요한 역할을 한다. 본 연구에서는 이러한 시스템 하에서, 추천 항목에 대해 사용자가 부여할 평가등급을 예측하는 방법을 제시하였다. 기존 연구에서는 주로 과거 평가등급의 평균을 기준으로 예측하였으나, 이는 평가등급 분포를 고려하지 않아 예측 오류 발생 가능성이 크므로, 평가등급의 범위를 근거로 새로운 예측 공식을 제안하였다. 제안 방법은 피어슨 상관도나 코사인 유사도를 이용하였을 때 모두 기존 방법보다 우수한 성능을 보였다. 향후 과제로서 제안 방법을 다른 유사도와 접목하였을 때의 성능 평가와 Book-crossing 데이터 외에 MovieLens, Jester 등의 데이터로 실험하여 그 성능을 조사 연구할 것이다.

참 고 문 헌

- [1] Arotaritei, D. & Mitra, S. (2004). Web mining: a survey in the fuzzy framework. *Fuzzy Sets and Systems*, 148(1), 5 - 19.
- [2] Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, 17(6), 734 - 749.
- [3] Shahabi, C. & Chen, Y.-S. (2003). Web information personalization: challenges and approaches. *3rd International Workshop on Databases in Networked Information Systems*. 5-15.
- [4] Su, X. & Khoshgoftaar, T.M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*.
- [5] Chen, L. & Sycara, K. (1998). WebMate: a personal agent for browsing and searching. *2nd International Conference on Autonomous Agents*, 132-139.
- [6] Liu, F., Yu, C., & Meng, W. (2004). Personalized web search for improving retrieval effectiveness, *IEEE Trans. Knowl. Data Eng.*, 16(1), 28-40.
- [7] Goldberg, D., Nichols, D., Oki, B.M., & Terry, D. (1992). Using collaborative filtering to weave an Information tapestry. *Comm. ACM*, 35(12), 61-70.
- [8] Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., & Riedl, J. (!997). GroupLens: applying collaborative filtering to usenet news. *Comm. ACM*, 40(3), 77-87.
- [9] Hill, W., Stead, L., Rosenstein, M., & Furnas, G. (1995). Recommending and evaluating choices in a virtual community of use. *Proc. Human Factors in Computing Systems*.
- [10] Terveen, L., Hill, W., Amento, B., McDonald, D., & Creter, J. (1997). PHOAKS: a system for sharing recommendations. *Comm.*

ACM, 40(3), 59-62.

- [11] Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: a constant time collaborative filtering algorithm. *Information Retrieval Journal*, 4(2), 133-151.
- [12] Hofmann, T. (2003). Collaborative filtering via gaussian probabilistic latent semantic analysis. *Proc. 26th Ann. Int'l ACM SIGIR Conf.*
- [13] Marlin, B. (2003). Modeling user rating profiles for collaborative filtering. *Proc. 17th Ann. Conf. Neural Information Processing Systems.*
- [14] Wang, T. & Ren, Y. (2009). Research on personalized recommendation based on web usage mining using collaborative filtering technique. *WSEAS Transactions on Information Science and Applications*, 6(1), 62-72.
- [15] Jeong, B., Lee, J., & Cho, H. (2010). Improving memory-based collaborative filtering via similarity updating and prediction modulation. *Information Sciences*, 180(5), 602-612.
- [16] Ahn, H. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, *Information Sciences*, 178(1), 37 - 51.
- [17] Ziegler, C.-N., McNee, S.M., Konstan, J.A., & Lausen, G. (2005) Improving recommendation lists through topic diversification. *Proc. the 14th International World Wide Web Conference*, 22 - 32.
- [18] Sarwar, B.M., Konstan, J.A., Borchers, A., Herlocker, J., Miller, B., & Riedl, J. Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. *Proc. the 1998 ACM Conference on Computer Supported Cooperative Work*, 345 - 354.



이수정

1985 이화여자대학교
과학교육과 (이학사)
1990 미국 Texas A&M 대학교
컴퓨터공학과 (석사)

1994 미국 Texas A&M 대학교 컴퓨터공학과
(박사)

1994~1998 삼성전자 통신개발실 선임연구원

1998~현재 경인교육대학교 컴퓨터교육과 교수

관심분야: 컴퓨터교육, 추천시스템, 웹마이닝

E-Mail: sjlee@gin.ac.kr