

Time-aware Item-based Collaborative Filtering with Similarity Integration

Soojung Lee*

*Professor, Dept. of Computer Education, Gyeongin National University of Education, Anyang, Korea

[Abstract]

In the era of information overload on the Internet, the recommendation system, which is an indispensable function, is a service that recommends products that a user may prefer, and has been successfully provided in various commercial sites. Recently, studies to reflect the rating time of items to improve the performance of collaborative filtering, a representative recommendation technique, are active. The core idea of these studies is to generate the recommendation list by giving an exponentially lower weight to the items rated in the past. However, this has a disadvantage in that a time function is uniformly applied to all items without considering changes in users' preferences according to the characteristics of the items. In this study, we propose a time-aware collaborative filtering technique from a completely different point of view by developing a new similarity measure that integrates the change in similarity values between items over time into a weighted sum. As a result of the experiment, the prediction performance and recommendation performance of the proposed method were significantly superior to the existing representative time aware methods and traditional methods.

▶ **Key words:** Collaborative Filtering, Item-based Collaborative Filtering, Time-aware Recommender System, Similarity Measure

[요 약]

인터넷 상의 정보 과부하 시대에 필수 불가결한 기능인 추천 시스템은 사용자가 선호할 만한 상품을 추천하는 서비스로서 여러 상업용 사이트에서 성공적으로 제공되고 있다. 최근 대표적인 추천 기법인 협력 필터링의 성능 개선을 위하여 항목 평가 시간을 반영하려는 연구가 활발하다. 이 연구들의 핵심 아이디어는 과거에 평가한 항목일수록 기하급수적으로 낮은 가중치를 두어 추천 결과를 산출하는 것이다. 그러나 이는 항목의 특성에 따른 사용자들의 선호도 변화를 고려하지 않고 모든 항목들에 대하여 일률적인 시간 함수를 적용한다는 단점을 가진다. 본 연구는 시간에 따른 항목 간의 유사도값 변화를 가중합으로 통합하는 새로운 유사도 척도를 개발함으로써 기존과 전혀 다른 관점의 시간 인지 협력 필터링 기법을 제안한다. 실험 결과, 제안 방법의 예측 성능과 추천 성능은 기존의 대표적 시간 인지 방법과 전통적 방법들에 비해 월등하게 우수하였다.

▶ **주제어:** 협력 필터링, 항목 기반 협력 필터링, 시각 인지 추천 시스템, 유사도 척도

-
- First Author: Soojung Lee, Corresponding Author: Soojung Lee
 - Soojung Lee (sjlee@gin.ac.kr), Dept. of Computer Education, Gyeongin National University of Education
 - Received: 2022. 06. 02, Revised: 2022. 07. 11, Accepted: 2022. 07. 12.

I. Introduction

현재 일상 생활에서 사람들은 뉴스, 음악, 여행 등의 다양한 소재에 대한 정보를 얻기 위해 다른 사람들의 경험이나 리뷰 등을 참조한다. 이러한 관점을 구현한 추천 시스템은 서적, 기사, 영화, 식당, 음악, 웹 쇼핑 등 매우 많은 영역에서 가장 흥미 있고 가치 있는 정보를 제공하기 위해 개발되었다. 추천 리스트를 형성하기 위해 시스템의 다른 사용자들의 평가 정보를 활용하는 방식을 협력 필터링(collaborative filtering, CF)이라고 한다[1][2].

CF 방식은 구현의 용이성 및 효율성으로 인해 상업용 시스템에 널리 적용되어 왔는데, 대표적으로 아마존, 넷플릭스, 반즈앤노블 등이 그 예이다. 이 방식의 기본 원리는 두 사용자의 항목 평가 행태 또는 선호도가 과거에 유사하였다면, 새로운 항목들에 대한 두 사용자의 향후 평가도 유사할 것으로 가정한다. CF 시스템은 직접 또는 간접적으로 입력된 사용자들의 평가정보를 활용하여 사용자들 간의 유사도를 측정한다. 간접 정보란 소비 행태, 클릭 횟수, 방문 횟수 등을 말한다[1][3].

평가 이력이 유사한 사용자들의 선호 항목들로부터 추천 결과를 산출해야 하기 때문에, 유사도의 정확한 측정은 CF 시스템의 성능을 좌우하는 매우 중요한 요소이다[2]. 관련 분야의 연구자들에 의해 다양한 척도들이 개발되었는데, 피어슨 상관 계수(Pearson correlation coefficient), 코사인 유사도(cosine-based similarity), 유클리디안 유사도(Euclidean distance-based similarity), 조정 코사인 유사도(Adjusted cosine-based similarity) 등은 대표적인 척도들이다[4].

학계 및 상업계에서의 CF 방식에 대한 인기에도 불구하고, 성능을 저해하는 여러 단점과 그에 대한 해결책들이 연구되었다. 데이터 희소성 문제(data sparsity problem)와 시스템 확장성(system scalability)이 주요 이슈로서 대두되었고, 그밖에 gray sheep, shilling attack 등도 해결해야 할 문제이다. 특히, 데이터 희소성 문제는 콜드 스타트와 신규 사용자 및 신규 항목 문제를 포함하고 있으며, 유사도 측정에 매우 큰 영향을 끼치므로 관련 학계에서 가장 주목하였던 문제들 중 하나이다[1][2].

정확한 유사도 측정을 위하여 항목 평가치 자체 뿐만 아니라 여러 가지 관련 요인, 예를 들어, 두 사용자의 공통평가항목수, 평가치의 분포 정보 등을 추가로 접목하여 새로운 유사도 척도를 개발하려는 노력이 진행되었다[5]. 사회적 관계망을 활용하여 보다 많은 평가정보를 획득하려는 시도 또한 발표되었다[2].

최근 인공 신경망의 발달로 인해 CF 시스템에도 이러한 기술을 적용하여 사용자들 간, 또는 사용자와 항목들 간의 복합적인 관계를 파악하여 성능을 개선하려는 연구가 진행되었다. 다양한 구조의 신경망, 즉, RBM, Autoencoder, CNN 등의 활용을 통한 CF 알고리즘의 개발이 그 예이다 [6]. 또 다른 접근 방식으로서 상황 인지(Context aware) 추천 시스템에 대한 연구가 지속적으로 진행되었는데, 사용자의 평가 시점을 반영한 시각 인지(Time-aware), 사용자의 위치를 파악하여 여행 가이드 등 추천에 활용하는 위치 인지(Location aware) 등이 포함된다[2][5].

본 연구는 새로운 관점의 평가 시각 인지 기반 협력 필터링 방법을 제안한다. 관련 연구 결과의 주된 방법은 각 사용자의 평가 시점이 과거일수록 기하급수적으로 감소하는 가중치를 부여하여 해당 항목이 추천될 가능성을 낮추는 방식을 취하였다. 이와는 달리 제안 방법은 주기적으로 항목 간의 유사도를 산출하고, 각 시점별 유사도값에 서로 다른 가중치를 부여함으로써 과거 평가치의 영향력을 감소시킨다. 성능 평가를 위하여, 다양한 시간 주기 및 가중치의 변화에 따른 예측 정확도와 정밀 정확도를 측정하였고, 결과적으로 제안 방법의 우수성을 입증하였다.

논문의 구성은 다음과 같다. 2절에서는 관련 지식 및 기존 연구 결과들을 소개한다. 3절에서는 제안 방법을 설명하고 4절에서 성능 실험 결과를 제시하며, 5절에서 논문의 결론을 맺는다.

II. Background

이웃 사용자들의 평가 데이터를 기반으로 추천 리스트를 결정하는 방식인 메모리 기반 CF 시스템은 크게 사용자 기반(user-based)과 항목 기반(item-based)으로 나뉜다[4]. 두 방식 모두 시스템의 현 사용자가 미평가한 항목들에 대해 추천 여부 또는 평가 예측치를 결정하기 위한 것이다.

관련 연구에서는 주로 사용자 항목 매트릭스(user-item matrix)를 활용하여 구체적인 알고리즘을 제시한다. 표 1은 평가치 1부터 5까지의 범위를 가진 시스템에서 5명의 사용자와 7개의 항목을 포함한 매트릭스 예시이다. u_1 이 현 사용자라고 할 때, 시스템에서는 u_1 의 미평가항목인 i_1, i_4, i_6 각각에 대한 예측치를 산출해야 한다.

Table 1. Illustration of an user-item matrix

	i1	i2	i3	i4	i5	i6	i7
u1	-	3	4	-	4	-	3
u2	1	2	-	-	2	4	2
u3	-	-	4	4	3	-	-
u4	3	2	-	-	3	-	-
u5	-	-	-	-	2	2	-

우선 미평가항목 i1에 대하여, 사용자 기반의 CF 시스템에서는 현 사용자 u1과 유사한 평가이력의 사용자들을 구한 후, 이들 이웃 사용자들의 i1에 대한 평가치를 통합하여 예측치를 구한다. 이 때 유사도값이 클수록 예측치에 대한 가중 영향력이 커진다.

실제 시스템에서는 대개 사용자수가 항목수를 훨씬 초과하며 대부분의 사용자는 극히 일부의 항목들에 대해서만 평가하므로, 사용자-항목 매트릭스의 셀은 대부분 공백이다. 따라서 사용자 간 유사도 측정값의 신뢰도가 저하될 수밖에 없다. 이러한 데이터 희소성 문제는 항목 간의 유사도 산출을 기반으로 하는 항목 기반 CF 방식에서 다소 완화된다. 이는 각 항목에 대한 평가치 밀도가 각 사용자에게 대한 그것보다 대체로 크기 때문이다. 표 1의 예에서, i1과 유사한 항목들은, u2와 u4가 모두 평가치를 부여한 i2와 i5이며, 이들에 대한 u1의 평가치는 각각 3과 4이므로 이들과 유사도값을 기초로 하여 예측치를 산출한다.

위 방식들은 모든 평가치를 시간적 관점에서 대등하게 취급하였으나, 시간 인지 CF에서는 과거 평가치에 대한 중요도를 낮게 책정한다. 초창기의 연구로서 이후 다른 관련 연구에서 흔히 인용된 [7]에서는 항목들을 그룹으로 나누어 각기 다른 감쇠율을 적용하였고, 그 성능 결과를 전통적 항목 기반 CF와 비교하였다. He와 Wu는 각 평가 시간대를 구간으로 나누어, 각 구간마다 피어슨 상관도를 산출하였고, 적은 평가개수의 구간에 대하여 낮은 가중치를 부여함으로써 산출된 유사도값의 신뢰도를 유지하는 방법을 취하였다[8]. 한편 이들과 유사한 방법으로서, [9]에서는 과거 평가치에 대하여 로지스틱 함수의 시간 가중치 함수를 적용하였다.

평가 시간을 반영한 최근의 연구 결과로서 Sun과 Dong은 사용자의 흥미 변화에 따른 예측 정확도를 개선하기 위하여 클러스터링과 시간 영향 지수 매트릭스를 도입하였다[10]. Ding 외 4인의 연구에서는 클라우드 서비스의 질 향상을 위하여 시간 특성과 데이터 희소성을 고려한 유사도를 채택한 후 ARIMA 모델을 적용하여 QoS 값을 예측하였다[11]. Li와 Han은 콘텐츠 기반과 협력필터링 기반 각각의 단점을 보완한 하이브리드 추천 방법을 제시

하고 시간 영향 지수 개념과 시간 인지의 사용자 선호 모델을 개발하였다[12]. 한편 시간 요소는 논문 추천에 있어서 매우 중요한 이슈이므로 이를 포함한 논문의 다차원 특성을 통합한 신경망 모델이 Lu 외 3인에 의해 제안되었다[13]. Wan 외 2인은 전통적인 사용자 기반과 항목 기반 CF 알고리즘에다가 시간 인지 잠재 요인 모델을 접목하여 새로운 추천 알고리즘을 제안하여 기존 모델들의 단점을 해결하려 하였다[14].

이상과 같이 시간 인지 요소는 주로 전통적인 유사도 기반의 협력 필터링 시스템에서 고려되었고, 과거 평가치에 대한 가중치 감쇠율의 적용이 전형적인 주된 방법이었다. 본 연구에서는 이와는 다른 새로운 접근 방식으로 시간 요소를 반영한 CF 알고리즘을 제안하며 다음 절에 상세 기술한다.

III. Proposed Methodology

1. Motivation and Idea

제안 방법의 개발 동기는 다음과 같다. 현 사용자 u의 미평가 항목 x와 y에 대하여, 만약 열 명의 이웃 사용자들이 부여한 x의 평가치 평균이 3.5이고, 두 명의 이웃 사용자들은 y에 대하여 평균 3의 평가치를 부여하였다고 하자. 이 항목들에 대한 평가 시간을 고려하지 않는다면 x가 y보다 u를 위한 추천 리스트에 포함될 가능성이 더 크다. 그러나 만약 y에 대한 평가시간이 더 최근이라면, 전형적인 시간 인지 협력 필터링 시스템은 y에 대한 예측치를 더욱 높게 산출할 수 있다. 이는 물론 시간에 따른 가중 함수값에 의해 좌우된다. 그러나, 두 항목에 대한 평균 평가치와 평가자 수를 살펴볼 때 x를 추천하는 것이 더욱 타당한 것으로 간주될 수 있다. 항목의 특성에 따른 사용자의 선호도 변화가 시간에 따라 미미한 경우 특히 그러하다.

위와 같은 논의 결과로서, 본 연구에서는 시간에 따라 항목 평가치에 대해 변화하는 가중치를 채택하지 않고, 유사도값에 대한 가중치 변화를 적용한다. 즉, 주기적으로 유사도값을 구한 후 최신의 유사도일수록 높은 가중치를 부여하여 통합 유사도를 산출한다. 또한, 사용자의 평가항목수는 전체 평가항목수에 비해 대체로 매우 작으므로 두 사용자 간에 산출된 유사도값의 신뢰도가 저하될 수밖에 없으므로, 이를 보완하기 위하여 항목 기반의 협력필터링 방식을 도입한다.

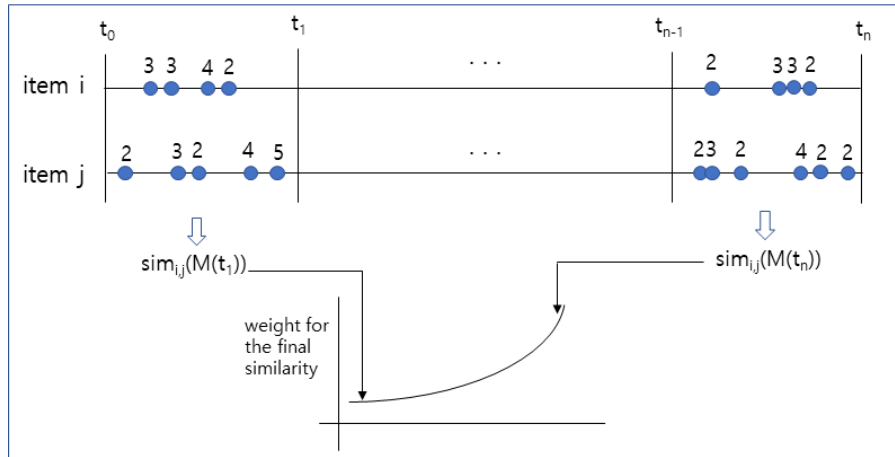


Fig. 1. Illustration of the proposed similarity integration

2. Formulation

평가 시각 $t_0 < t_1 < t_2 < \dots < t_n$ 에 대하여, 사용자-항목 매트릭스 $M(t_k)$ 는 $t_{k-1}+1$ 시각부터 t_k 시각까지의 모든 항목에 부여된 평가치를 포함한다고 하자. 즉, $\cap M(t_k) = \emptyset$ 이다. 각 매트릭스 내 평가치들을 사용하여 임의의 두 항목 간의 유사도를 산출한다. 이 때 관련 연구에서 가장 활용도가 높은 것으로 알려진 피어슨 상관을 유사도로 정의한다.

각 $M(t_k)$, $k=1, \dots, n$ 에 대하여, 두 항목 i 와 j 간의 유사도 산출 공식은 다음과 같다.

$$sim_{i,j}(M(t_k)) = \frac{\sum_{u \in U_i \cap U_j} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U_i \cap U_j} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U_i \cap U_j} (r_{u,j} - \bar{r}_j)^2}} \dots \dots \dots (1)$$

위 식에서 U_i 는 $M(t_k)$ 내에 i 를 평가한 모든 사용자들의 집합, $r_{u,i}$ 는 $M(t_k)$ 에 포함된 i 에 대한 사용자 u 의 평가치,

그리고 $\bar{r}_i = \sum_{u \in U_i \cap U_j} r_{u,i}$ 이다.

각 시간 구역에 대하여 유사도값을 산출한 후에, 두 항목 i 와 j 간의 최종 유사도는 각 유사도값의 가중합으로서 아래와 같이 정의한다. 최근의 유사도값일수록 기하급수적으로 가중합으로써 최종 유사도에 대한 영향력의 크기를 증가시켰다.

$$sim_{i,j} = \frac{\sum_{k=1}^n e^{\lambda k} \cdot sim_{i,j}(M(t_k))}{\sum_{k=1}^n e^{\lambda k}}, \lambda > 0 \dots \dots (2)$$

그림 1은 위에서 기술한 유사도 산출의 일례이다. 두 개의 항목 i 와 j 에 대하여 가로축은 시간대를 나타내며, 시스템 사용자들의 평가 시점을 점으로 표시하였고 평가치를 함께 나타내었다. 그림에서는 각 시간 구간 M 에 대하여 유사도를 산출한 연후에, 최종 유사도를 식 (2)에 의해 산출하기 위하여 가중치값을 결정하는 예시도 함께 제시하였다.

3. Rating Prediction

일반적인 메모리 기반 협력필터링 시스템에서 사용자의 미평가 항목에 대한 예측치를 산출하는 방법은 크게 두 종류이다. 첫째는 항목을 평가한 다른 사용자들의 평가치의 가중합이고, 두 번째는 단순 가중 평균인데 이는 주로 항목 기반 예측치 산출을 위해 활용된다. 본 연구에서는 첫째 방법을 이용하여 사용자 u 의 미평가항목 x 에 대한 예측치, $\hat{r}_{u,x}$, 를 아래와 같이 산출한다. 이 식에서 \bar{r}_u 는 u 의 모든 평가치의 평균이며, j 는 u 가 평가치를 부여한 시스템 내 항목이고 그 평가치는 $r_{u,j}$ 로 표기하였다.

$$\hat{r}_{u,x} = \bar{r}_u + \frac{\sum_j (r_{u,j} - \bar{r}_u) \cdot sim_{x,j}}{\sum_j |sim_{x,j}|} \dots \dots \dots (3)$$

IV. Performance Experiments

1. Design of Experiments

협력 필터링 알고리즘의 성능 실험은 대개 관련 연구에서 널리 활용하는 공개 데이터셋을 통하여 이루어진다. 본 연구에서는 시간 데이터를 포함한 MovieLens를 활용하였다 (<http://www.movielens.org/>). 이 셋에서 제공하는 평가

데이터 포맷은 사용자ID, 영화ID, 평가치, 타임스탬프로 구성된다. 데이터 규모는 100M까지 제공되나, 실험용 PC의 메모리 성능을 감안하여 2000명의 사용자가 3952개의 영화에 대한 평가치로 국한하여 실험하였다. 결과적으로 전체 평가 개수는 339,541 건이므로 각 사용자 당 평균 평가개수는 약 170개이다. 평가 범위는 1~5의 정수 형태이다.

제안 알고리즘의 성능을 객관적으로 살펴보기 위하여, 다음과 같은 여러 방법들의 성능 결과와 비교하였다.

- COR_UB: 피어슨 상관의 유사도를 이용한 사용자 기반 협력 필터링 방법
- COR_IB: 피어슨 상관의 유사도를 이용한 항목 기반 협력 필터링 방법
- PROP_UB: 제안 방법의 통합 유사도를 사용자 기반으로 구현한 협력 필터링 방법
- PROP_IB: 제안 방법의 통합 유사도를 항목 기반으로 구현한 협력 필터링 방법
- DL: 피어슨 상관의 유사도를 항목 기반으로 구현한 시간 인지 기반의 협력 필터링 방법[7]

2. Performance Metrics

추천 시스템의 성능 평가는 크게 두 가지로 이루어지는데, 즉, 예측 정확도와 추천 정확도이다. 본 연구의 실험에서는 대표적인 예측 정확도 척도로서 가장 널리 활용되는 MAE(Mean Absolute Error)와 RMSE(Root Mean Squared Error)를 도입하였다. MAE는 시스템에서 추정된 미평가항목의 예측치가 얼마나 실제치에 근접하는지를 측정하며, 구체적인 정의는 아래와 같다.

$$MAE = \frac{1}{n} \sum_u \sum_x |r_{u,x} - \hat{r}_{u,x}| \quad \dots\dots\dots(4)$$

MAE는 각 평가치 차이를 동일한 가중으로 취급한 결과이다. 그러나, RMSE는 예측치와 실제치의 차이가 큰 경우에는 예측 성능이 더욱 저하된 것으로 간주하기 위하여, 다음과 같이 정의한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_u \sum_x (r_{u,x} - \hat{r}_{u,x})^2} \quad \dots\dots\dots(5)$$

마지막으로, 시스템에서 추천한 항목에 대한 사용자의 만족 여부를 파악하기 위하여, 평균 정밀도 (mean average precision, MAP)를 측정하였다[15]. 실험 데이터셋의 평가 범위가 1부터 5인 점을 고려하여, 4 이상의 평가치를 부여한 항목은 사용자가 만족하는 것으로 간주하였다.

3. Effect of Parameters

3.1 The Effect of Decay Rate

감쇠율(decay rate)은 식 (2)에서 각 시간대별 유사도값의 비중을 좌우하는 파라미터이다. 그림 2는 PROP_IB에서 감쇠율 λ값의 변화에 따른 MAE 성능을 측정된 결과이다.

작은 λ값일 경우에 예측 성능의 차이는 미미한 결과를 보였는데, 대체로 λ값이 커질수록 성능은 더욱 향상되었다. 그러나, 향상 정도는 매우 작은 것으로 나타났다. λ가 커지면 가장 최근의 유사도값의 영향을 더욱 크게 반영하는 것이므로, 앞 절에서 설명한 제안 방법의 동기와 부합됨을 확인하였다. 이 결과에 따라, 본 연구에서는 λ=1.5로 정하였다.

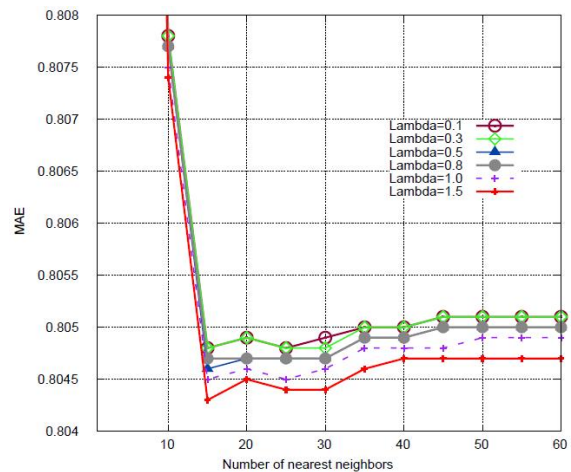


Fig. 2. MAE results of PROP_IB with varying decay rate

3.2 The Effect of Time Interval

제안 방법은 시간 구간으로 평가 데이터를 구분하여 유사도값을 산출한 후 가중합의 최종 유사도를 사용하므로, 모두 몇 개의 구간으로 분할하는가가 성능을 좌우하는 요소가 된다. 본 실험에서는 데이터셋의 전체 기간인 약 34개월을 5, 6, 7, 8, 9, 10, 12개의 구간(TI)으로 나누어 그림 3과 같은 예측 성능 결과를 얻었다.

대체로 TI가 작거나 큰 경우에 저조한 성능을 보였고, 중간 정도의 값인 7 또는 8의 경우에 더 우수한 결과를 보였다. TI가 작으면 각 부분데이터 크기가 작아지므로, 산출된 유사도값의 신뢰가 떨어지고, 그 반대의 경우엔 시간 인지 효과가 충분히 반영되지 않으므로 성능이 저하되는 것으로 판단된다. 따라서, TI=8로 설정하여 이후의 실험을 진행하였다.

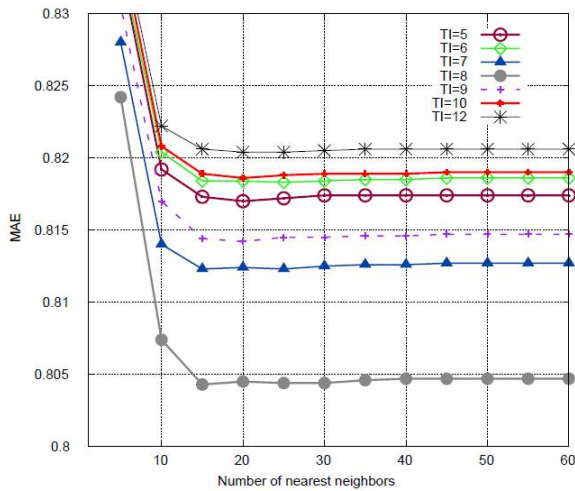


Fig. 3. MAE results of PROP_IB with varying time interval

4. Performance Results

그림 4,5,6은 IV.1절에서 언급한 다섯 가지 방법의 성능을 비교한 결과이다. MAE와 RMSE에 대하여 알고리즘들의 비교 양상이 유사하게 나타났는데, 그 이유는 두 척도의 정의가 유사하기 때문이다. 두 결과에서 모두 제안한 시간 인지 방법이 기존의 피어슨 유사도 기반의 방법을 월등하게 능가하는 성능을 나타냈다. 그러나, PROP_UB가 전통적인 항목 기반 방법인 COR_IB의 성능을 뛰어넘지 못하므로, 사용자 기반의 협력필터링 방식의 한계를 알 수 있다. 즉, 앞 절에서 언급했듯이, 각 사용자 당 170개의 평가개수는 전체 항목의 약 4.3%를 차지하므로, 사용자 간의 공통평가 항목개수에 의존하여 유사도를 산출하는 사용자 기반 협력 필터링 방식의 성능은 이러한 환경의 영향을 받는다.

DL은 항목 기반의 방법이므로 예측 정확도 성능은 물론 COR_UB보다 우수하였다. 또한, COR_IB보다 우수한 결과를 나타낸 것으로 보아, 각 항목에 대해 지수 함수의 평가시간 가중치를 부여한 아이디어가 유효함을 알 수 있다. 그러나, MAE 결과에서는 COR_IB의 성능보다 약간 우수하고, PROP_IB보다 훨씬 저하된 성능을 보이므로, DL 방법의 한계를 드러냈다. 결과적으로, 유사도의 시간 변화를 고려한 제안 방법은 예측 성능 면에서 더 우수함을 확인하였다.

시스템의 추천 리스트에 대한 사용자 만족도를 살펴 보기 위하여 그림 6에 MAP 결과를 제시하였다. 추천 항목수가 증가함에 따라 정의에 의해 정밀도는 감소되었는데, 다만 COR_UB는 나머지 네 방법보다 추천 리스트의 크기가 60일 때 현격하게 저하된 성능을 보였다. 또한 DL의 정밀도는 기존의 항목 기반 방식을 넘어서지 못하는 거의 대등한 결과를 보인데 반해, PROP_IB는 이들보다 약간 우수한 성능을 보였다. 종합적으로, 제안 방법의 항목 간 유사도

값의 시간 인지 알고리즘은 기존의 항목 시간 반영의 알고리즘보다 예측 성능과 추천 성능에 대하여 우수한 결과를 보였으며, 특히 예측 성능 면에서 더욱 월등한 것으로 나타났다.

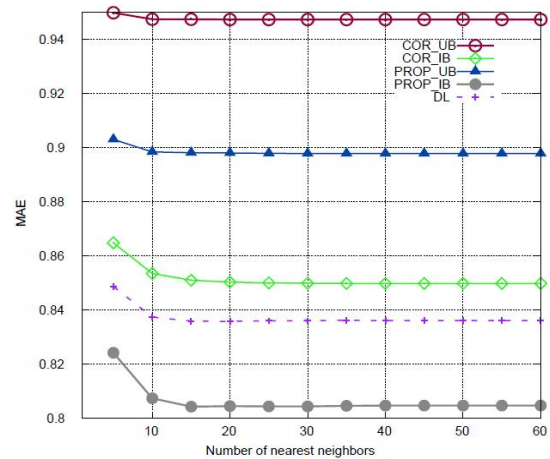


Fig. 4. MAE results of different methods

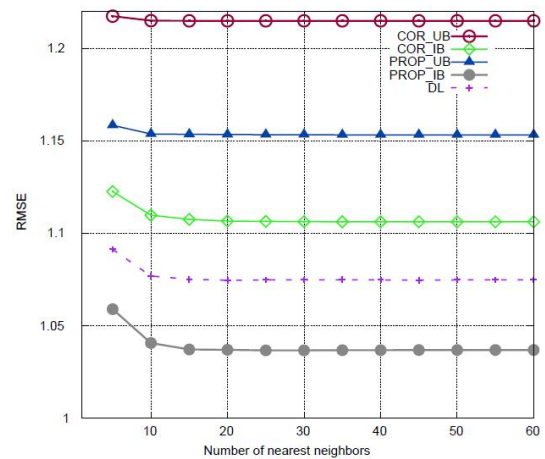


Fig. 5. RMSE results of different methods

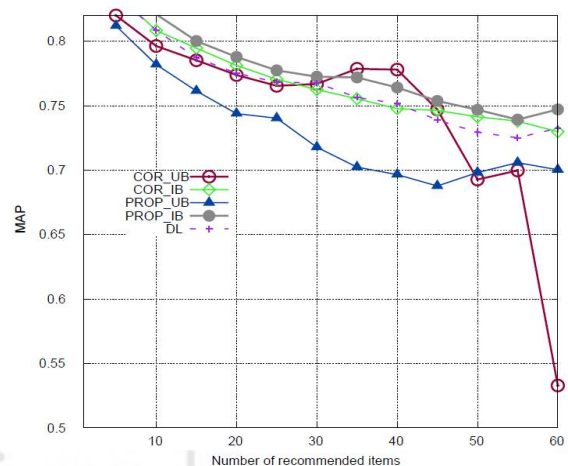


Fig. 6. MAP results of different methods

V. Conclusions

본 연구에서는 평가 시각 인지 기반의 협력 필터링 시스템을 위한 새로운 유사도 척도를 제안하였다. 기존 방법들의 핵심 아이디어가 항목 평가 시각이 과거일수록 해당 평가치에 가장 감쇠율을 적용한 것과는 달리, 주기적으로 항목 간 유사도를 측정하여 유사도값의 시간에 따른 변화를 고려한 통합 유사도 척도를 개발하였다. 성능 평가를 위하여 예측 성능과 추천 성능을 측정 후 기존 방법들과 비교한 결과, 특히 예측 성능에서 제안 방법은 매우 월등한 결과를 나타냈다. 본 연구는 항목 기반의 추천 시스템이므로 시스템의 새로운 사용자들도 기존 사용자들과 동일하게 시간 인지 기반의 항목 추천을 받아볼 수 있다는 잇점이 있다.

추천 시스템에서 시간 인지 개념의 도입 방법과 관련된 연구는 활발히 이루어져 왔으나, 평가 시각에 좌우하는 가중치 함수의 적용이 보편적이다. 본 연구에서도 기존의 일반적인 가중치 함수인 지수함수를 적용하여 다양한 파라미터값의 변화에 따른 성능 결과를 살펴 보았으나, 향후 시간 함수의 종류나 평가된 항목의 특성에 따른 차별화된 적용 방법 등에 대한 연구가 추가 진행되어야 할 과제이다. 또한, 연구용 개방 데이터셋으로 널리 활용되는 MovieLens 외에 평가 시각 정보를 제공하는 데이터셋은 거의 전무하므로, 본 연구의 아이디어를 일반적으로 검증하기에 어려움이 있었으나, 다른 특성을 가진 데이터셋이 향후 제공된다면 연구 내용을 심화시키기에 매우 큰 도움이 될 것으로 생각된다.

REFERENCES

- [1] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutierrez, "Recommender Systems Survey," *Knowledge-based Systems*, Vol. 46, pp. 109-132, 2013. DOI: 10.1016/j.knsys.2013.03.012
- [2] M. Jalili, S. Ahmadian, M. Izadi, P. Moradi, and M. Salehi, "Evaluating Collaborative Filtering Recommender Algorithms: A Survey," *IEEE Access*, Vol. 6, pp. 74003-74024, 2018. DOI: 10.1109/ACCESS.2018.2883742
- [3] K. Choi, D. Yoo, G. Kim, and Y. Suh, "A Hybrid Online-product Recommendation System: Combining Implicit Rating-based Collaborative Filtering and Sequential Pattern Analysis," *Electronic Commerce Research and Applications*, Vol. 11, No. 4, pp. 309-317, 2012. DOI: 10.1016/j.elerap.2012.02.004.
- [4] C. Desrosiers and G. Karypis, "A Comprehensive Survey of Neighborhood-based Recommendation Methods," *Recommender Systems Handbook*, pp. 107-144, Springer, Berlin, Germany, 2011. DOI: 10.1007/978-1-4899-7637-6_2
- [5] J. Bobadilla, A. Hernando, F. Ortega, and J. Bernal, "A Framework for Collaborative Filtering Recommender Systems," *Expert Systems with Applications*, Vol. 38, No. 12, pp. 14609-14623, 2011. DOI: 10.1016/j.eswa.2011.05.021
- [6] Z. Batmaz, A. Yurekli, A. Bilge, and C. Kaleli, "A review on deep learning for recommender systems: challenges and remedies," *Artificial Intelligence Review*, vol. 52, no. 1, pp. 1-37, 2019.
- [7] Y. Ding and X. Li, "Time Weight Collaborative Filtering," *Fourteenth ACM international Conference on Information and Knowledge Management*, pp. 485-492, 2005. DOI: 10.1145/1099554.1099689
- [8] L. He and F. Wu, "A Time-Context-Based Collaborative Filtering Algorithm," *IEEE International Conference on Granular Computing*, pp. 209-213, 2009. DOI: 10.1109/GRC.2009.5255130
- [9] Y. Huai-Zhen and L. Lei, "An Enhanced Collaborative Filtering Algorithm Based on Time Weight," *International Symposium on Information Engineering and Electronic Commerce*, pp. 262-265, 2009. DOI: 10.1109/IEEC.2009.61
- [10] B. Sun and L. Dong, "Dynamic Model Adaptive to User Interest Drift based on Cluster and Nearest Neighbors." *IEEE Access* 5, 1682-1691 (2017) DOI: 10.1109/ACCESS.2017.2669243
- [11] S. Ding, Y. Li, D. Wu, Y. Zhang, and S. Yang, "Time-aware Cloud Service Recommendation using Similarity-enhanced Collaborative Filtering and ARIMA Model," *Decision Support Systems*, Volume 107, 2018, Pages 103-115, ISSN 0167-9236, <https://doi.org/10.1016/j.dss.2017.12.012>.
- [12] H. Li and D. Han, "A Time-aware Hybrid Recommendation Scheme Combining Content-based and Collaborative Filtering." *Frontiers of Computer Science*. 15, 154613 (2021). <https://doi.org/10.1007/s11704-020-0028-7>
- [13] Y. Lu, Y. He, Y. Cai, and Z. Peng, "Time-aware Neural Collaborative Filtering with Multi-dimensional Features on Academic Paper Recommendation May 2021 DOI:10.1109/CS-CWD49262.2021.9437673 Conference: 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)
- [14] Y. Wan, Y. Chen, and C. Yan, (2021). "An Integrated Time-Aware Collaborative Filtering Algorithm." *Knowledge Management in Organizations*. pp. 369-379, 2021. https://doi.org/10.1007/978-3-030-81635-3_30
- [15] Y. Deldjoo, M.F. Dacrema, M.G. Constantin, et al., "Movie genome: alleviating new item cold start in movie recommendation," *User Modeling and User-Adapted Interaction* 29, 291-343 (2019) <https://doi.org/10.1007/s11257-019-09221-y>

Authors



Soojung Lee received the B.S. degree in Mathematics Education from Ewha Woman's University, Korea in 1985. She received M.S. and Ph.D. degrees in Computer Science from Texas A&M University in 1990 and 1994,

respectively. Dr. Lee joined the faculty of the Department of Computer Education at Gyeongin National University of Education, Gyunggi-do, Korea, in 1998, as a professor. She is interested in recommender systems, information filtering, data mining techniques, and computer education.