

Integration of Similarity Values Reflecting Rating Time for Collaborative Filtering

Soojung Lee*

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

[Abstract]

As a representative technique of recommender systems, collaborative filtering has been successfully in service through many commercial and academic systems. This technique recommends items highly rated by similar neighbor users, based on similarity of ratings on common items rated by two users. Recently research on time-aware recommender systems has been conducted, which attempts to improve system performance by reflecting user rating time of items. However, the decay rate uniform to past ratings has a risk of lowering the rating prediction performance of the system. This study proposes a rating time-aware similarity measure between users, which is a novel approach different from previous ones. The proposed approach considers changes of similarity value over time, not item rating time. In order to evaluate performance of the proposed method, experiments using various parameter values and types of time change functions are conducted, resulting in improving prediction accuracy of existing traditional similarity measures significantly.

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

[요 약]

협력 필터링은 추천 시스템의 대표적인 기법으로서 많은 상업 및 학계 시스템에서 성공적으로 구현되어 서비스되고 있다. 이 기법은 두 사용자 간의 공통 평가 항목에 대한 평가치의 유사성을 기반으로 유사한 이웃 사용자들이 높은 평가치를 부여한 항목들을 추천한다. 최근 사용자들의 항목 평가 시각을 반영하여 시스템 성능을 향상시키려는 시각 인지 추천 시스템 연구가 진행되고 있다. 그러나, 과거 평가치에 대한 일률적인 감쇠율은 시스템의 평가치 예측 성능을 저하시킬 우려가 있다. 본 연구에서는 기존과 다른 접근 방식으로서 평가 시각 인지 기반의 사용자 간 유사도 척도를 제안한다. 이 방법은 항목 평가 시각이 아닌 유사도값의 시간에 따른 변화를 고려한다. 제안 방법의 성능 평가를 위해 다양한 파라미터값과 시간 변화 함수 종류에 대하여 실험 평가를 진행하였으며, 기존의 전통적인 유사도 척도들의 예측 성능을 크게 향상시키는 결과를 나타냈다.

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

-
- First Author: Soojung Lee, Corresponding Author: Soojung Lee
 - *Soojung Lee (sjlee@gin.ac.kr), Dept. of Computer Education, Gyeongin National University of Education
 - Received: 2021. 12. 13, Revised: 2022. 01. 05, Accepted: 2022. 01. 05.

I. Introduction

현재 인터넷 시대의 정보량은 급격히 증가하므로, 사용자들은 원하는 정보를 처리 및 추출하는데 막대한 시간과 비용을 지불하는 추세이다. 따라서, 사용자에게 적합한 정보를 선별하여 제공하는 추천 시스템(recommender system)은 인터넷을 통한 각종 서비스계의 필수 불가결한 요소이다. 사용자에게 적합한 정보의 식별은 직접 또는 간접적으로 이루어지는데, 사용자로부터 평가 데이터를 요청하거나 또는 사용자의 소비 행태, 클릭 및 다운로드 횟수, 웹페이지 방문 여부 등을 통해서 이루어진다[1][2].

추천 시스템의 구현을 위해 다양한 필터링 방법이 개발되었는데, 협력 필터링(collaborative filtering), 내용 기반 필터링(content-based filtering), 인구통계학적 필터링(demographic filtering)을 비롯하여 이들의 혼합 방식인 하이브리드 필터링(hybrid filtering)이 대표적이다. 이후에는 사회 연결망 기반 필터링(social network based filtering), 지식 기반 필터링(knowledge-based filtering) 등이 대두되어 기존 방식의 단점을 개선하려고 시도하였다[2]. 최근에는 사용자의 평가 시각을 고려한 시각 인지(Time-aware)나 위치 인지(Location aware) 등의 상황 인지(Context aware) 추천 시스템에 대한 연구가 활발하다. 또한 유전자 알고리즘(Genetic algorithm)이나 신경망(Neural network)을 주로 사용하는 생체 영감(Bio-inspired) 접근 방식에 대한 관심도가 높아지고 있다[3][4].

협력 필터링(CF)은 다른 기법들과는 달리 사용자의 직, 간접적 평가 데이터만을 필요로 하기 때문에 구현이 용이하여 매우 활발히 연구되어 왔으며 넷플릭스, 아마존 등의 실제 시스템에 적용되어 서비스되고 있다. CF를 위한 가장 널리 알려진 알고리즘은 kNN(k Nearest Neighbors)으로서 현 사용자와 가장 유사한 평가 데이터를 가진 이웃 사용자들을 구한 후, 추천 리스트를 결정하기 위하여 현 사용자가 미평가한 항목들의 평가치를 예측한다[3]. 이웃 사용자들과의 유사도 값을 우선순위로 하여 그들의 평가치를 통합하여 예측치를 산출한다.

이와 같은 유사 사용자 또는 유사 항목 산출을 토대로 하는 방식을 이웃 기반(neighborhood-based) CF라고 하는데, 추천 시스템에서 가장 널리 사용되는 기법이다[2][3]. 또한, 사용자의 특성이나 평가 정보로부터 모델을 구축하여 미평가 항목의 평가치를 예측하는데 활용하는 모델 기반의 CF(model-based CF)와는 달리 이웃 기반 CF는 정확도를 유지하면서도 간단한 구현 방식으로 인해

상업계와 학계에서 널리 활용되어 왔다. 이웃 기반 CF에서 유사도 값의 측정은 예측 정확도를 결정하는 매우 핵심적인 요소이므로 다양한 척도들이 개발되었다. 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[5][6].

본 연구는 평가 시각 인지 기반의 사용자 간 유사도 척도를 제안한다. 대부분의 기존 연구에서는 사용자들의 평가치에 대하여 감쇠율(decay rate)을 적용함으로써, 과거의 평가치일수록 예측치 산출에 미치는 영향을 감소시키는 방식을 취하여, 항목의 특성에 무관하게 시간에 따라 동일한 감쇠율을 적용하였다. 이와는 달리, 제안 방법은 항목 평가 시각이 아닌 유사도 값의 시간에 따른 변화를 고려한다. 제안 방법에서 도입한 최적의 파라미터 값과 시간 변화 함수를 구하기 위하여 다양한 조건에서의 실험 평가를 진행하였으며, 그 결과 기존의 여러 유사도 척도들의 예측 성능을 크게 향상시킴을 확인하였다.

논문의 구성은 다음과 같다. 2절에서는 시각 인지 협력 필터링에 관련된 기존 연구 결과들을 소개한다. 3절에서는 제안 방법을 설명하고 4절에서 성능 실험 결과를 제시하며, 5절에서 논문의 결론을 맺는다.

II. Related Works

사용자 기반의 협력 필터링(user-based collaborative filtering)의 근본 원리는 시스템의 현 사용자와 유사한 평가 이력의 사용자들이 선호하였던 항목들을 추천하는 것이다. 두 사용자 간의 유사도를 측정하기 위한 방법으로, 그들이 공통으로 평가했던 항목들의 평가치에 대하여 다양한 유사도 척도를 적용하는 방식이 대표적이다. 현 사용자가 미평가한 항목을 추천할지의 여부는 해당 항목에 대한 평가 예측치를 산출한 다음 예측 값이 기준보다 높으면 추천 대상으로 선정한다. 예측치는 유사한 이웃 사용자들의 해당 항목에 대한 평가치로부터 산출한다.

이와 같은 전통적인 방법은 이웃 사용자가 항목을 평가한 시점을 고려하지 않으므로, 예측 정확도가 저하될 우려가 있다. 따라서, 다양한 상황 인지 CF(context-aware CF) 기법들 중 하나인 평가 시각 인지 CF(time-aware CF)는 이웃 사용자의 과거 평가치에 대해 기하급수적으로 감소하는 가중치를 부여하여 최종 예측치를 산출하는 방

식을 취한다[7][8].

Ding과 Li는 항목군별로 다른 감쇠율을 적용하기 위해 항목 클러스터링을 실시하고 전통적인 항목 기반의 CF 알고리즘과 성능을 비교하였다[9]. H.-Zhen과 Lei는 과거 평가치에 시간 가중치를 곱한 후에 피어슨 상관도로 유사도를 계산하였고, 가중치 함수로서 로지스틱을 도입하였다 [10]. He와 Wu는 각 평가 시간 구간별로 사용자 간의 피어슨 상관도를 산출하였으며, 각 구간에서의 평가개수가 적을 때 예측치 가중을 감소하여 희소 데이터에 따른 유사도 값의 신뢰도 저하를 보완하였다[11].

한편 사용자가 방문하였던 문서의 태그 정보를 추가적으로 감안하여 사용자의 태그 선호도를 반영하는 방법을 개발한 연구가 Zheng과 Li에 의해 발표되었는데, 이 연구 결과의 특징은 다른 연구들과 마찬가지로 과거 정보일수록 낮은 가중치를 두지만, 사용자의 선호 정보를 추가함으로써 과거 정보일지라도 선호 문서일 경우에는 좀 더 높은 가중치를 부여할 수 있다는 점이다[12]. 이밖에도 [13]에서는 좀 더 다양한 정보를 통합하여 유사도 척도를 제안하였는데, 공통 평가항목 수 및 이들의 평가치 차이, 평가 시각의 차이, 평가 순서의 차이 등의 정보를 도입하였고, 이들 요소들의 각 비중은 파라미터 값의 변화 실험을 통해 정하였다. 이와 유사한 방법으로서 [14]는 평가치 차이, 평가 시각 차이, 그리고 평가 순서 차이의 합을 정규화한 값으로 산출하는 새로운 유사도 척도를 제안하였고, 공통 평가치가 없는 두 사용자를 위한 이행성의 유사도 값 (transitive similarity) 산출 방식을 고안하였다.

Lai와 Deng은 사용자 기반이 아닌 항목 기반의 CF로서 각 항목이 평가된 시각을 고려하여 지수적으로 감소하는 가중치를 부여하였다[15]. 결과적으로 이 방식에서도 오래된 평가치는 작은 비중으로 취급한다는 점에서 기존과 동일하다고 볼 수 있다. 이 같은 시각 인지 CF 기반의 추천 시스템에 관한 전반적인 연구 결과물들과 평가 척도에 관하여 조사한 논문이 [16]에 의해 발표되었는데, 훈련 데이터와 시험 데이터로의 분할 문제, 데이터 분할과 연계된 비율 기반과 시각 기반의 데이터 크기 문제, 다양한 성능 평가 척도와 여러 유형의 데이터 셋을 활용한 시각 인지 알고리즘의 성능 결과가 제시되었다. 최근의 연구 결과로서 Xu 외 4인은 사용자의 확신도와 시각 인지를 반영한 조정 코사인 유사도(adjusted cosine similarity)를 제안하였다[17]. [18]에서도 시간 함수를 고려함과 동시에 텍스트 리뷰를 활용함으로써 데이터 희소성 문제에 대한 해결책을 제시하였다. 결론적으로 대부분의 시각 인지 기반의

추천 시스템에서는 항목에 대한 과거 평가치에 대하여 항목 특성 또는 기타 기준에 대한 고려 없이 동일한 감쇠율을 적용하여 낮은 가중치를 부여하는 방식을 취하였다.

III. Proposed Methodology

본 연구에서는 항목 평가 시각을 반영하는 사용자 간 유사도 척도를 제안한다. 제안 방법의 개발 동기는 다음과 같다. 만약 한 사용자가 디즈니 만화 영화 A를 선호하여 t 시각에 5.0/5.0 만점의 평가치를 부여하였다고 하자. 이 사용자가 t+α 시각에 같은 영화에 대해 이를테면 4.0의 평가치, 그리고 t+2α 시각에는 2.0의 평가치를 부여할 것으로 가정하는 것은 다소 비합리적이다. A 영화는 과거 영화로서 비추천 대상은 될 수 있지만, 그 평가치를 시간 경과에 따라 기하급수적으로 감소시켜 참조하는 기존 연구 방법들은 재고의 여지가 있다는 가정 하에, 본 연구에서는 평가치가 아닌 유사도 값을 대상으로 시간에 따른 변화를 반영하였다. 즉, 매 시간 주기별로 사용자 간의 유사도 값을 산출하고, 과거 유사도 값에 대하여는 낮은 가중치를 부여함으로써 최종 유사도 값을 산출한다. 구체적인 절차는 아래와 같다.

1. 전체 평가 데이터를 평가 시각을 기준으로 하여 N개의 부분집합으로 나눈다. 즉, 전체 데이터는 R₁UR₂U...UR_N와 동일하며, 임의의 항목 i에 대한 평가 시각을 t(i)라고 할 때, i∈R_n, j∈R_{n+1}에 대하여 t(i) < t(j)이다. 최적의 N 값은 실험에 의해 결정할 파라미터이다.
2. 각 부분 평가 데이터 집합 R_n, n=1, ..., N에 대하여, 임의의 두 사용자 u와 v 간의 유사도를 구한다. 아래 식에서 유사도 측정을 위해 피어슨 상관도를 활용하였으며, \bar{r}_u 는 R_n에 속한 사용자 u의 모든 평가 데이터의 평균을 나타낸다.

$$sim_{u,v}(R_n) = \frac{\sum_i (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_i (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_i (r_{v,i} - \bar{r}_v)^2}} \dots\dots\dots(1)$$

$i \in R_n, r_{u,i} \neq NULL, r_{v,i} \neq NULL$

3. 2번 과정에서 획득한 sim_{u,v}(R₁), ..., sim_{u,v}(R_N)을 통합한 최종 유사도 값을 아래와 같이 산출한다.

$$sim_{u,v} = \frac{\sum_{n=1}^N w(R_n) \cdot sim_{u,v}(R_n)}{\sum_{n=1}^N w(R_n)} \dots\dots\dots(2)$$

각 부분 데이터 집합 R_n 에 대한 유사도 값 $sim_{u,v}(R_n)$ 에 해당하는 가중치는 기하급수적 또는 선형적 함수를 활용하여 부여하되, 이는 실험에 의해 결정할 파라미터이다. 즉, $w(R_n) = e^{\lambda n}$ 또는 λn 이며, λ 값이 클수록 최근 평가 데이터를 활용한 유사도 값의 영향이 커진다. λ 는 실험 파라미터이다.

4. 사용자 u 와 모든 다른 사용자와의 유사도 값이 결정되면, u 가 미평가한 항목 x 에 대한 평가 예측치를 다음과 같이 산출한다[3].

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

5. 사용자 u 를 위한 항목 추천 리스트는 4번 절차에서 산정한 예측치의 내림차순으로 구성한다.

위와 같은 제안 방법에서 각 부분 평가 데이터 집합에 대해 유사도를 산출해야 하므로, 모두 N 번의 계산 횟수가 소요된다. 이는 기존에 단 한 번의 유사도 산출 과정이 필요했던 것에 비해, 단점이라 할 수 있다. 그러나, 부분 데이터 집합은 평가 시각을 기준으로 나누어지므로, 매 시간 주기마다 최신 데이터에 대한 단 한 번의 유사도를 산출하면 되므로, 결과적으로 기존 방법과 마찬가지로 복잡도를 가진다. 또한, 제안 방법은 전체 데이터가 아닌 특정 시간 범위 내의 부분 데이터만을 활용하여 유사도를 측정하므로, 기존보다 취급하는 평가 데이터 양이 현저히 감소하므로 유사도 산출 시간 또한 현저히 단축될 수 있다. 다만, 부분 데이터 양은 대개 희소하므로 이로 인하여 산출된 유사도 값의 신뢰도가 저하될 우려가 있다. 이러한 문제는 본 연구의 범위를 벗어난 것으로서, 데이터 희소성 문제를 다른 기존의 여러 개발 방법들을 활용하여 해결해야 할 향후 과제이다. 제안 방법의 2번 절차에서 피어슨 상관도를 도입하였는데, 이는 대표적인 CF 알고리즘으로서, 관련 연구 분야에서 널리 사용되어 온 유사도 산출 방법이다. 이 밖에도 전통적인 유사도 산출 방법, 또는 최근의 이웃 기반의 CF 방법들을 대신 활용하여도 제안 방법의 근본 취지에서 벗어나지 않으므로 무방하다.

IV. Performance Experiments

1. Design of Experiments

제안 방법의 성능 실험을 위하여 평가 시각 정보를 포함한 데이터 셋이 필요하므로, 기존 연구에서 활발히 사용되어 왔던 MovieLens(<http://www.movielens.org/>)를 실험

데이터로 선정하였다. 이 데이터 셋은 6040명의 사용자가 3952개의 영화에 대해 2003년부터 약 34개월 동안 부여한 1부터 5까지의 정수 평가치를 포함한다. 대부분의 기존 연구의 실험 방식대로, 전체 데이터 중 80%는 사용자 간의 유사도 값 산출을 위한 훈련 데이터로 사용하였고, 나머지 20%는 훈련의 결과로서 시스템의 성능을 측정하기 위한 용도로 사용하였다.

본 논문에서는 미평가 항목에 대한 예측치 산출 방법을 제시하였으므로, 성능 평가를 위하여 CF 알고리즘에서 사용하는 대표적인 예측 정확도의 평가 척도를 도입하였다. MAE(Mean Absolute Error)가 대표적인데, 이는 각 사용자의 실제 평가치와 시스템의 예측 평가치의 차이의 평균으로 정의하며 아래와 같다.

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

이에 더하여, 두 평가치의 차이를 더욱 증폭하여 반영하는 또 다른 척도로서 RMSE(Root Mean Squared Error)를 도입하였는데, 이 척도는 넷플릭스 대회에서 활용되어 널리 알려졌다.

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

성능 비교 실험을 위하여 전통적인 유사도 척도와 대비하여 평가 시간을 반영한 척도가 얼마나 큰 성능 향상을 가져오는지 알아보았다. 따라서 대표적 기존 척도인 피어슨 상관도(COR, Pearson correlation), 코사인 유사도(COS, Cosine similarity), 평균자승차이(MSD, Mean squared differences)와 이들 각각에 본 연구의 시간 반영 알고리즘을 적용한 방법들을 TARS_COR, TARS_COS, TARS_MSD로 표기하여 비교하였다.

2. The Effect of Decay Rate

본 절에서는 제안 방법의 가중 함수에서 사용하는 decay rate인 λ 값에 따른 성능 변화를 파악하였다. 그림 1은 다양한 λ 값의 지수 함수 $e^{\lambda n}$ 를 가중치로 하였을 때의 MAE 성능 결과이다. λ 값이 커질수록 식 (2)에서 가장 최근의 유사도 값의 가중 함수 결과 값이 커진다. 즉, 최근의 유사 사용자의 평가치가 예측에 더욱 큰 영향을 미친다.

예상대로 $\lambda=5.0$ 인 경우 가장 좋은 성능을 나타냈고, 0.1일 때 가장 저조하였다. 특히, $\lambda=2.0$ 보다 클 때는 성능 향상의 정도가 미미한 것으로 볼 수 있다. 한편, 식 (2)에서 가중 함수를 선형으로 하여 실험한 결과에서는 λ 값의 영향이 그림 1의 결과보다 적었으며, 성능 결과 또한 그림 1의 최고 성능보다 저조한 것으로 확인되었다.

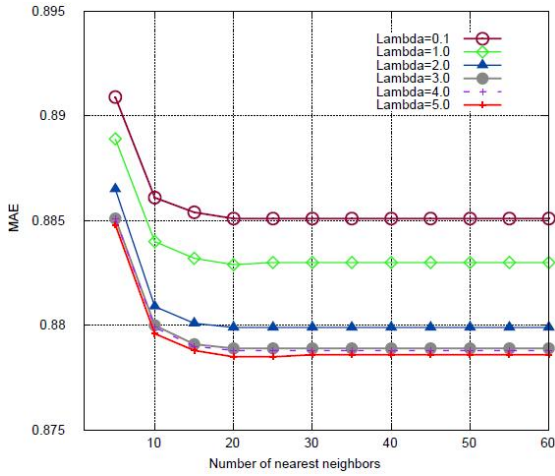


Fig. 1. Performance results with varying lambda

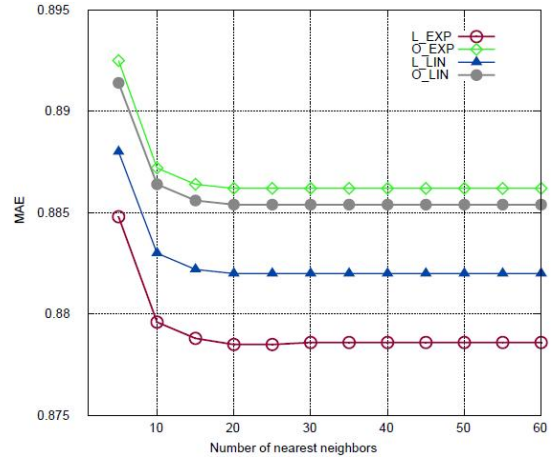


Fig. 2. Performance results with different weight functions

3. The Effect of Time Weight Types

시간의 변화에 따른 가중 함수는 두 사용자의 최종 유사도값에 영향을 미치는 매우 중요한 요소이다. 따라서 함수의 종류가 성능에 어떠한 영향을 미치는지 실험하였고, 그 결과를 그림 2에 제시하였다. 범례에 표기된 함수의 정의는 다음과 같다.

- L_EXP: 지수 함수이며, 최신 평가치일수록 큰 가중치. $w(R_n) = e^{\lambda n}$.
- L_LIN: 선형 함수이며, 최신 평가치일수록 큰 가중치. $w(R_n) = \lambda n$.
- O_EXP: 지수 함수이며, 과거 평가치일수록 큰 가중치. $w(R_n) = e^{-\lambda n}$.
- O_LIN: 선형 함수이며, 과거 평가치일수록 큰 가중치. $w(R_n) = -\lambda n$.

그림에서 보듯이, 과거 평가치를 반영하는 방식에 비해 최신 평가치를 반영하는 방법의 성능이 월등히 우수하였다. 또한, 과거 평가치 반영 방식은 함수 종류에 따른 영향이 상대적으로 크지 않았으며, 최근의 유사도 값을 기하급수적으로 반영하는 L_EXP가 가장 우수한 성능을 보였다. 따라서 앞 절의 실험 결과와 종합하여, 이후의 실험은 $\lambda = 5.0$ 의 지수 함수를 적용하여 계속 진행하였다.

4. The Effect of Time Interval

제안 방법은 전체 평가 데이터를 여러 부분 데이터로 분할하여 유사도를 측정한다. 부분 데이터 집합의 크기가 시스템 성능에 미치는 영향을 알아보기 위하여, 다양한 시간 구간(Time Interval, TI)에 대하여 MAE 결과를 산출하여 그림 3에 제시하였다. TI 값이 크면 부분 집합의 크기가 커지고, 사용자 간의 유사도 산출을 위한 데이터 량도 많아지므로 산출된 유사도 값의 신뢰도가 높아진다. 반면에, 긴 시간 구간 내의 모든 평가치가 동일한 가중치를 갖게 되므로, 시간 변화에 따른 사용자의 평가 행태 추이를 정밀하게 반영하진 못한다.

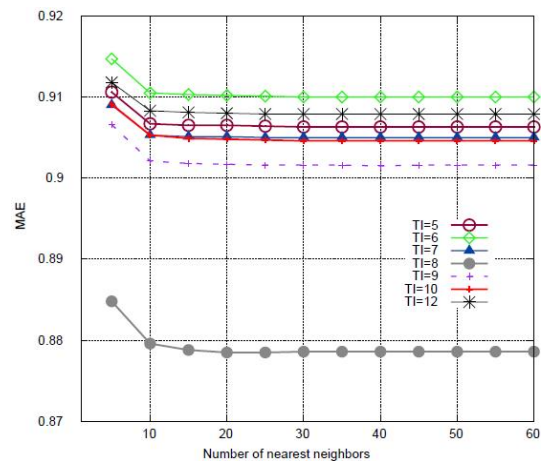


Fig. 3. Performance results with different time intervals

그림 3에서 TI=8인 경우에 월등하게 우수한 성능을 보였다. MovieLens의 평가 데이터는 대략 34개월 동안의 평가 이력을 제공하는데, TI=8은 하나의 부분 데이터 집합이 약 4.3 개월의 구간이 된다. 앞 절에서 언급한 논리대로, 구간이 너무 짧거나 길면 성능에 악영향을 미침을 알 수 있다.

5. Performance Comparison

제안 방법의 성능을 기존 방법들과 비교하기 위해, 이전 실험 결과를 토대로 하여 최고의 성능을 보였던 L_EXP 지수 가중 함수를 적용하고, $TI=8$, $\lambda=5.0$ 로 하였다.

그림 4와 5는 평가치를 참조할 이웃 사용자 수의 증가에 따른 MAE와 RMSE 결과를 나타낸다. 두 가지 결과는 비슷한 양상을 보였는데, COS는 매우 저조한 성능을 보였고 그 다음으로 COR 성능이 저조하였다. 이는 데이터 셋이 충분한 양의 평가 데이터를 포함하지 않기 때문으로 판단되며, 이러한 환경에서는 MSD가 더 유리함을 알 수 있다. 반면에 시간 변화에 따른 유사도 값을 적용한 제안 방법은 이들 기존 척도들의 성능을 크게 향상시켰는데, 특히 COS에 대해 가장 큰 효과를 보였다.

RMSE는 예측치가 실제치와 크게 다를 때 이를 반영하는 척도인데, 그림 4에서 TARS_COR와 TARS_MSD는 거의 같은 MAE 성능 결과를 보였으나, 그림 5에서 TARS_MSD의 RMSE 성능이 약간 더 우수한 것으로 나타났으므로, TARS_MSD가 전반적으로 실제치에 더욱 가까운 예측 결과를 산출함을 알 수 있다. 그러나, 제안 방법은 실험에 사용한 기존 척도들 중에서 COS에 적용하였을 때 가장 큰 예측 성능 개선 효과를 나타냈다.

이와 같은 예측 성능 결과는 COS에 대하여 약 7%, COR에 대하여 약 4.2%의 향상도를 보였으므로, [10]에서 제안한 방법이 기존 방법에 비해 약 3.5%의 성능 향상을 보인 것과 비교할 때 우수함을 알 수 있다. 또한, [17]의 제안 방법의 향상도는 COS와 대비하여 약 3.2%, COR과 대비하여 약 1.5%로 보고되었으므로, 본 연구의 제안 방법이 매우 우수함을 알 수 있다.

V. Conclusions

본 연구에서는 협력 필터링 시스템을 위한 새로운 평가 시각 인지 기반의 유사도 척도를 제안하였다. 제안 척도는 사용자의 항목 평가 시각에 따른 평가치 가중 감쇠율을 항목 특성에 무관하게 적용하는 기존 연구 방법을 개선하기 위하여, 사용자 간 유사도 값의 시간에 따른 변화를 고려하였다. 성능 평가를 위하여 다양한 파라미터 값과 시간 변화 함수 종류에 대하여 실험을 진행하였으며, 기존의 전통적인 유사도 척도들의 예측 성능을 크게 향상시키는 결과를 나타냈다. 특히, 평균자승차이의 유사도 척도를 활용하였을 때 가장 우수한 성능을 보였다.

시간에 따른 평가 데이터의 변화를 어떻게 반영할지의 연구는 최근 활발히 진행되고 있으나, 아직 많은 이슈가 논의되고 있는 분야이다. 향후 변화 대상의 종류, 즉, 항목 평가치 또는 유사도 값 등, 에 대한 분석과 확장 연구가 더욱 필요하다. 또한, 변화의 주기, 항목 특성에 따른 변화 함수 파악 등의 연구도 함께 진행되어야 할 부분이다.

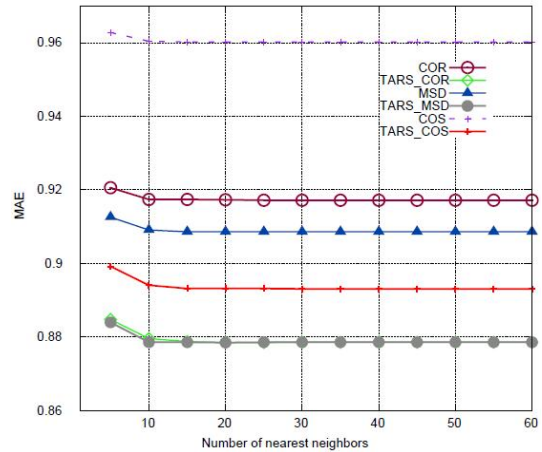


Fig. 4. MAE results of different similarity measures

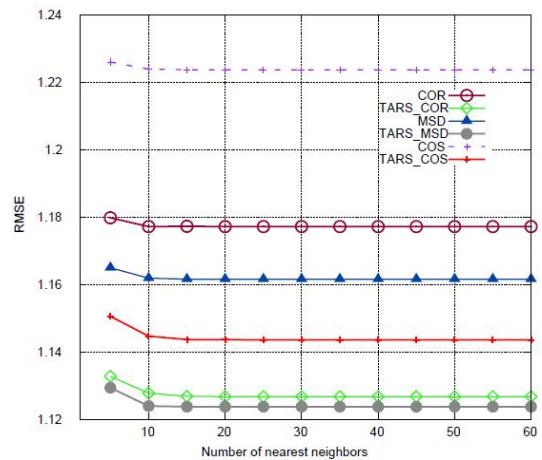


Fig. 5. RMSE results of different similarity measures

ACKNOWLEDGEMENT

This research was supported by the academic research grant of Gyeongin National University of Education in 2021.

REFERENCES

- [1] 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.
- [2] 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.knosys.2013.03.012
- [3] 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
- [4] 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
- [5] 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
- [6] H. Khojamli and J. Razmara, "Survey of Similarity Functions on Neighborhood-based Collaborative Filtering," *Expert Systems with Applications*, Vol. 185, 2021, Article Number 115482, DOI: 10.1016/j.eswa.2021.115482
- [7] A. Livne, E. S. Tov, A. Solomon, A. Elyasaf, B. Shapira, and L. Rokach, "Evolving Context-aware Recommender Systems with Users in Mind," *Expert Systems with Applications*, Vol. 189, 2022, Article Number 116042, DOI: 10.1016/j.eswa.2021.116042
- [8] N. Mohammadi and A. Rasoolzadegan, "A Two-stage Location-sensitive and User Preference-aware Recommendation System," *Expert Systems with Applications*, Vol. 191, 2022, Article Number 116188, DOI: 10.1016/j.eswa.2021.116188
- [9] 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
- [10] 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
- [11] 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
- [12] N. Zheng and Q. Li, "A Recommender System based on Tag and Time Information for Social Tagging Systems," *Expert Systems with Applications*, Vol. 38, No. 4, pp. 4575-4587, 2011. DOI: 10.1016/j.eswa.2010.09.131
- [13] A. Bahrehmand and R. Rafeh, "Proposing a New Metric for Collaborative Filtering," *Journal of Software Engineering and Applications*, Vol. 4, No. 7, pp. 411-416, 2011. DOI: 10.4236/jsea.2011.47047
- [14] R. Rafeh and A. Bahrehmand, "An Adaptive Approach to Dealing with Unstable Behaviour of Users in Collaborative Filtering Systems," *Journal of Information Science*, Vol. 38, No. 3, pp. 205-221, 2012. DOI: 10.1177/0165551512437517
- [15] W. Lai and H. Deng, "An Improved Collaborative Filtering Algorithm Adapting to User Interest Changes," *The 6th International Conference on New Trends in Information Science, Service Science and Data Mining*, pp. 598-602, 2012.
- [16] P. G. Campos and F. Diez and I. Cantador, "Time-aware Recommender Systems: A Comprehensive Survey and Analysis of Existing Evaluation Protocols," *User Modeling and User-Adapted Interaction*, Vol. 24, No. 1, pp. 67-119, 2014. DOI: 10.1007/s11257-012-9136-x
- [17] G. Xu, Z. Tang, C. Ma, Y. Liu, and M. Daneshmand, "A Collaborative Filtering Recommendation Algorithm Based on User Confidence and Time Context," *Journal of Electrical and Computer Engineering*, Vol. 2019, Article ID 7070487, 12 pages, DOI: 10.1155/2019/7070487
- [18] C. Wangwatcharakul and S. Wongthanavas, "A Novel Temporal Recommender System based on Multiple Transitions in User Preference Drift and Topic Review Evolution," *Expert Systems with Applications*, Vol. 185, 2021, Article Number 115626, DOI: 10.1016/j.eswa.2021.115626

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, Gyeonggi-do, Korea, in 1998, as a professor. She is interested in recommender systems, information filtering, data mining techniques, and computer education.