

<https://doi.org/10.7236/JIIBC.2021.21.4.183>
JIIBC 2021-4-26

추천 시스템을 위한 단계적 평가치 예측 방안

A Stepwise Rating Prediction Method for Recommender Systems

이수정*

Soojung Lee*

요약 협력 필터링 기반의 추천 시스템은 현재 다양한 분야의 상업용 시스템의 필수불가결한 기능으로서, 사용자들이 선호할만한 상품을 맞춤형으로 제공해 주는 유용한 서비스이다. 그러나, 사용자들의 평가 데이터가 불충분할 경우 선호 상품의 예측이 부정확할 우려가 크다. 본 연구에서는 이러한 단점을 해결하기 위하여 단계적으로 상품의 평가치를 예측하는 방안을 제시한다. 각 단계에 해당하는 예측 방법의 적용 조건을 만족하지 못할 경우 다음 단계의 방법을 적용한다. 제안 방법의 성능 평가를 위해, 공개 데이터셋을 활용한 실험을 진행하였으며, 제안 방법은 여러 전통적 유사도 척도를 도입한 협력 필터링 시스템의 예측 성능과 정밀도 성능을 크게 향상시켰고, 평가데이터 희소성 해결을 위한 기존 방식들의 성능을 능가하는 결과를 보였다.

Abstract Collaborative filtering based recommender systems are currently indispensable function of commercial systems in various fields, being a useful service by providing customized products that users will prefer. However, there is a high possibility that the prediction of preferable products is inaccurate, when the user's rating data are insufficient. In order to overcome this drawback, this study suggests a stepwise method for prediction of product ratings. If the application conditions of the prediction method corresponding to each step are not satisfied, the method of the next step is applied. To evaluate the performance of the proposed method, experiments using a public dataset are conducted. As a result, our method significantly improves prediction and precision performance of collaborative filtering systems employing various conventional similarity measures and outperforms performance of the previous methods for solving rating data sparsity.

Key Words : Collaborative filtering, Data sparsity, Rating prediction, Recommender system

*정회원, 경인교육대학교 컴퓨터교육과
접수일자 2021년 6월 29일, 수정완료 2021년 7월 29일
게재확정일자 2021년 8월 6일

Received: 29 June, 2021 / Revised: 29 July, 2021 /
Accepted: 6 August, 2021

*Corresponding Author: sjlee@ginue.ac.kr

Dept. of Computer Education, Gyeongin National University of
Education, Korea

I. 서 론

협력 필터링(collaborative filtering, CF)은 추천 시스템의 주요한 방식으로 많은 상업용 사이트에서 실제로 유용하게 사용되고 있다^{[1][2]}. 서적, 음악, 여행 등의 정보를 사용자 맞춤형으로 제공하여 정보 검색 시간을 절약할 수 있고 사용자의 선호 상품을 용이하게 구할 수 있도록 도움을 제공한다. 협력 필터링 관련 연구는 수십 년간 활발히 이루어져 왔으며^{[3][4]}, 최근 대두된 다양한 인공지능 알고리즘을 적용하여 사용자의 선호 특성을 파악하기 위한 연구도 진행되어 왔다^{[5][6]}.

이웃 기반(neighborhood-based) CF 알고리즘에서는 인접 이웃 사용자들의 상품 평가 이력을 활용하여 현 사용자가 미평가한 상품들의 선호도를 예측한다. 인접 이웃은 사용자들이 직접 또는 간접적으로 부여한 상품에 대한 평가치를 기반으로 현 사용자와 유사한 평가 이력의 사용자들로 정의한다^{[7][8]}. 인접 이웃이 누구인가에 따라서 미평가 상품의 예측치가 다르게 산출되므로, 인접 이웃을 결정하는 유사도 척도는 CF 기반 시스템의 성능을 좌우하는 매우 중요한 요소들 중 하나이다.

본 연구에서는 CF 시스템에서 평가치 예측의 본질적인 문제, 즉, 인접 이웃의 수가 적어서 상품의 평가치를 참조하기에 불충분한 경우의 해결책을 제안한다. 제안 방법의 특징은 사용자 기반과 항목 기반의 예측치 산출 방법을 단계적으로 적용하되, 각 단계의 방식을 적용하기 위한 예측치 산출 조건을 정의한다. 즉, 조건을 만족하는 경우에 해당 방식을 적용하여 보다 신뢰할만한 예측치를 산출할 수 있다. 제안 방법은 공개 데이터셋을 활용한 실험을 통해 성능을 측정하였으며, 여러 전통적 유사도 척도를 도입한 CF 시스템의 예측 성능과 정밀도 성능을 크게 향상시켰고, 평가데이터 희소성 해결을 위한 기존 방식들의 성능을 능가한 결과를 보였다.

논문의 구성은 다음과 같다. 2절에서는 관련 연구를 소개하고 3절에서는 본 연구의 새로운 평가치 예측 방법을 설명한다. 4절에서 성능 측정 실험 결과를 제시하며, 5절에서 논문의 결론을 맺는다.

II. 관련 연구 내용

1. 유사도 산출

협력 필터링 시스템에서 평가 이력에 대한 두 사용자 간 또는 두 항목 간의 유사도 산출 방법은 전통적으로 공

통의 평가치들을 활용한다. 두 사용자 간의 유사도는 그들이 공통으로 평가한 항목들의 평가치로부터, 두 항목 간의 유사도는 그들을 모두 평가한 사용자들의 평가치로부터 구한다. 대표적인 유사도 산출 방법으로서 상관도 기반의 피어슨 상관(Pearson correlation)이 널리 활용되어 왔다^[8]. 두 사용자 u 와 v 간의 유사도는 아래 식과 같이 산출된다.

$$COR(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}} \dots (1)$$

위 식에서 $r_{u,i}$ 는 사용자 u 의 항목 i 에 대한 평가치, \bar{r}_u 는 u 의 평가치 평균, $I_{u,v}$ 는 u 와 v 의 공통평가항목 집합이다. 이밖에도 상관도 기반의 유사도 산출 방법으로서 제한 피어슨 상관도(Constrained Pearson Correlation), 스피어만 순위 상관(Spearman rank correlation), 켄델 타우 상관(Kendall's tau correlation) 등이 포함된다^[9].

또다른 대표적 유사도 산출 방법으로서 코사인 유사도(cosine similarity)는 사용자들의 평가치를 벡터로 취급하며 다음과 같이 정의된다.

$$COS(u, v) = \frac{\sum_{i \in I_{u,v}} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I_{u,v}} r_{u,i}^2} \sqrt{\sum_{i \in I_{u,v}} r_{v,i}^2}} \dots \dots \dots (2)$$

조정된 코사인 유사도(adjusted cosine similarity)는 사용자의 서로 다른 평가등급 스케일을 반영하기 위하여 개발되었다^[7].

이밖에 평균자승차이(mean squared differences, MSD)는 공통평가항목 각각에 대해 두 평가치 차이의 제곱의 평균값으로서 아래와 같이 산출한다.

$$MSD(u, v) = \frac{1}{I_{u,v}} \sum_{i \in I_{u,v}} (r_{u,i} - r_{v,i})^2 \dots \dots \dots (3)$$

각 항목에 대한 평가치 차이가 작을수록 유사도는 커지므로, 두 사용자간 유사도는 $1-MSD$ 로 정의한다^[8].

위 방법들의 가장 큰 문제점은 사용자의 평가 데이터가 희소한 경우에 유사도 값의 산출이 아예 불가능하거나, 또는 산출된 유사도 값의 신뢰도가 떨어진다는 점이다. 따라서, 두 사용자의 공통평가항목수를 반영하여 유사도 산출의 정확도를 높이려는 시도가 연구되었는데, 이 중 대표적인 방법으로서 자카드 계수가 포함된다. 이 계수는 두 사용자가 평가한 모든 항목들 중에서 공통으로 평가한 항목수의 비율을 말한다^{[10][11]}. 구체적으로, 사

용자 u 가 평가한 항목들의 집합을 I_u 라고 할 때, 자카드 계수는 다음과 같이 정의된다.

$$Jaccard(u, v) = \frac{|I_u \cap I_v|}{|I_u \cup I_v|} \dots\dots\dots(4)$$

자카드 계수는 주로 전통적인 유사도 척도와 결합하는 방식으로 흔히 사용되어 왔는데, 특히 희소한 평가 데이터 환경에서 기존 유사도 척도의 단점을 보완하여 성능 개선을 가져온다는 다수의 연구가 발표되었다^[12]. Bobadilla 외 2인은 MSD와 자카드 계수를 결합한 유사도 척도를 제안하고, 개선된 성능 실험 결과를 보고하였다^[13]. 이 연구의 척도는 이후 많은 연구자들에 의해 비교 분석 대상이었다^[14]. Saranya 외 2인은 피어슨 상관도와 결합했을 때의 성능 개선 효과를 분석하였다^[9]. 자카드 계수 외에 유사한 성격으로서 공통 평가 항목만을 고려하는 트라이앵글 유사도를 활용한 연구가 발표되었는데, Iftikhar 외 4인은 트라이앵글 유사도에 기반한 상품 추천 방법을 제시하였고^[15], Sun 외 6인의 연구에서는 트라이앵글 유사도를 자카드 계수와 통합하여 예측 오류를 개선하려 하였다^[11].

2. 평가치 예측 방법

모든 사용자 간 또는 항목 간의 유사도값을 산출한 후에, 협력 필터링 시스템은 현 사용자가 미평가한 항목에 대해 평가치를 예측하고, 예측 평가치가 높은 순으로 해당 항목들을 추천한다. 이는 평가 이력이 유사한 인접 사용자나 항목 간의 유사도 값을 토대로 하며, 향후 다른 항목에 대한 평가치도 유사하게 부여할 것이라는 가정을 근거로 한다. 어떠한 사용자나 항목들이 인접이웃 집합에 속하는지에 따라 평가 예측치가 달라지기 때문에, 인접이웃을 결정하는 유사도 산출 방식은 매우 중요하며 이에 따라 시스템의 성능이 좌우된다.

평가치를 예측하는 방법은 주로 사용자 기반과 항목 기반으로 나뉘는데 주로 가중합(weighted sum)으로서 예측치를 결정한다. 특히, 사용자 기반의 예측방법은 각 사용자의 평가치 평균을 반영하는 방법이 널리 활용되며, 항목 기반의 방법으로는 단순 가중합(simple weighted average)이 개발되었다^[8].

III. 단계적 평가치 예측 방법

본 연구에서는 사용자 기반과 항목 기반의 예측 방법

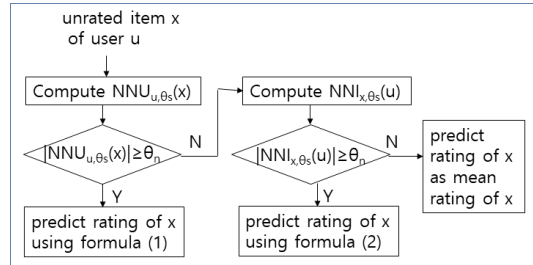


그림 1. 현 사용자 u 가 미평가한 항목 x 에 대한 평가치 예측 절차

Fig. 1. Procedure of rating prediction for unrated item x of the target user u

표 1. 제안 방법에서 사용하는 기호와 의미

Table 1. Description of symbols used by the proposed method

기호	의미
$r_{u,x}$	사용자 u 의 항목 x 에 대한 평가치
$\hat{r}_{u,x}$	사용자 u 의 미평가 항목 x 에 대한 평가 예측치
\bar{r}_u	사용자 u 의 평가치 평균
NNU_u	사용자 u 의 인접 사용자 집합
NNI_x	항목 x 의 인접 항목 집합
θ_s	유사도 값의 하한선
θ_n	인접 이웃수의 하한선
$NNU_{u, \theta_s}(x)$	사용자 u 와의 유사도값이 θ_s 이상이며 항목 x 를 평가했던 인접 사용자 집합
$NNI_{x, \theta_s}(u)$	항목 x 와의 유사도값이 θ_s 이상이며 사용자 u 가 평가했던 인접 항목 집합

을 단계적으로 적용하는 새로운 방식을 제안한다. 제안 방식의 장점은 기존에 고려하지 않았던 사용자 또는 항목의 인접이웃수의 부족에 따른 희소성 문제를 해결하여 예측치의 정확성을 향상시키려 한 점이다. 그림 1은 제안 방법의 예측치 산정 절차를 나타내며 그림에서 사용한 기호들의 의미는 표 1에 제시하였다. 현 사용자 u 가 미평가한 항목 x 에 대한 평가치를 예측하기 위하여, 우선 유사도 값이 θ_s 보다 크며 x 에 대한 평가 이력이 있는 인접 사용자들을 구한다. 이들의 숫자가 θ_n 보다 크다면 신뢰할 만한 예측치를 산출할 충분한 크기로 판단한다. 따라서 사용자 기반의 방법인 아래 식에 의거하여 예측치를 산출한다^[8].

$$\hat{r}_{u,x} = \bar{r}_u + \frac{\sum_v sim(u,v)(r_{v,x} - \bar{r}_v)}{\sum_v |sim(u,v)|} \dots\dots\dots(5)$$

위 식에서 $v \in NNU_{u, \theta_s}(x)$ 이다.

만약 위에서 언급한 인접 사용자들의 수가 θ_n 보다 작을 경우엔 항목 기반의 예측 방법을 활용한다. 즉, x 와

표 2. 성능 비교 대상 방법

Table 2. Methods for Performance Comparison

기호	유사도 산출 방법	예측 평가치 산출 방법
COR	피어슨 상관도 (Pearson correlation)	사용자 기반 방법 (공식 (5))
COS	코사인 유사도 (Cosine similarity)	
MSD	평균 자승 차이 (Mean squared differences)	
COR_P	피어슨 상관도 (Pearson correlation)	제안 방법
COS_P	코사인 유사도 (Cosine similarity)	
MSD_P	평균 자승 차이 (Mean squared differences)	
JAC	자카드 계수 (Jaccard index)	사용자 기반 방법 (공식 (5))
JMSD	JAC * (1-MSD)	사용자 기반 방법 (공식 (5))

유사한 인접 항목들 중에서 유사도 값이 θ_s 보다 크며 u 가 평가한 이력이 있는 인접 항목들을 구한다. 이들의 숫자가 θ_n 보다 크다면 신뢰할 만한 예측치를 산출할 충분한 크기라고 판단하여 항목 기반의 방법인 아래 식으로서 예측치를 산출한다.

$$\hat{r}_{u,x} = \bar{r}_x + \frac{\sum_j sim(x,j) \times (r_{u,j} - \bar{r}_j)}{\sum_j |sim(x,j)|} \dots\dots\dots(6)$$

위 식에서 $j \in NN_{x,\theta_s}(u)$ 이다.

만약 위의 인접 항목들의 숫자도 θ_n 보다 작을 경우에는 x 에 대한 평가치들의 평균값으로 예측치를 정한다. 제안 방법에서 사용하는 두 하한선 θ_s 과 θ_n 의 적정값은 실험을 통하여 구한다.

IV. 성능 평가 실험

1. 실험 환경

제안 방법의 성능 평가를 위하여, 관련 연구에서 널리 활용되며 공개 데이터셋인 Jester를 채택하여 실험하였다^{[7][8]}. 이 데이터셋은 1000명의 사용자의 100개의 농담 항목에 대한 -10부터 10까지의 실수형 평가치를 포함한다. 실험을 위하여 유사도 산출을 위한 훈련 데이터 집합을 80%, 성능 평가를 위한 시험 데이터 집합을 20%로 전체 데이터셋을 분할하였다. 시스템의 추천 항목에 대하여 평가치가 4 이상이면 적합한 것으로 판정하여 성능 결과를 산출하였다.

2. 성능 평가 척도

협력 필터링 시스템의 성능은 미평가치에 대한 예측

정확도, 항목 추천 리스트에 대한 적합도, 추천 항목들의 순위 성능 등으로 평가된다. 이를 위하여 관련 연구에서 널리 활용되는 MAE(Mean Absolute Error), RMSE (Root Mean Squared Error), nDCG(Normalized Discounted Cumulative Gain), MAP((Mean Average Precision)을 도입하였다.

3. 성능 비교 대상 방법

전통적인 유사도 척도들을 사용자 기반 예측치 산출 방법(공식 5)과 제안 방법을 각각 적용하여 성능 결과를 산출하였으며, 데이터 희소성을 해결하기 위하여 제안된 기존 연구 방법인 자카드 계수와 JMSD^[13]를 추가하여 성능 비교하였다. 표 2에 실험 대상 방법들을 정리 요약하였다.

4. 실험 결과

가. 파라미터값의 영향

그림 2는 제안 방법에서 사용하는 파라미터들인 θ_s 와 θ_n 값의 변화가 평균 절대 오차(MAE) 성능에 미치는 영향을 실험한 결과이다. 대체적으로 θ_s 값이 커질수록 성능은 떨어지며, 특히 $\theta_s=0.5$ 일 때 급격히 저하되었다. 이는 θ_s 값보다 작은 유사도의 이웃들은 평가치를 참조할 인접 이웃에서 제외되었기 때문에 당연한 결과이다. 그러나, $\theta_s=0.2$ 인 경우가 $\theta_s=0.1$ 일 때보다 더 좋은 결과를 보이는데, θ_s 값이 너무 작으면 유효하지 않은 인접이웃들이 포함되기 때문이다.

그림 2에서 θ_n 의 변화에 따른 성능도 제시하였는데, 값이 커질수록 우수한 성능을 나타냈다. 이는 미평가 항목의 평가치 예측을 위해 참조할 인접 이웃들의 수가 많아지기 때문이다. 그러나, θ_n 의 크기가 너무 크다면 조건을 만족할 인접이웃들을 구하기 어려워지며, 따라서

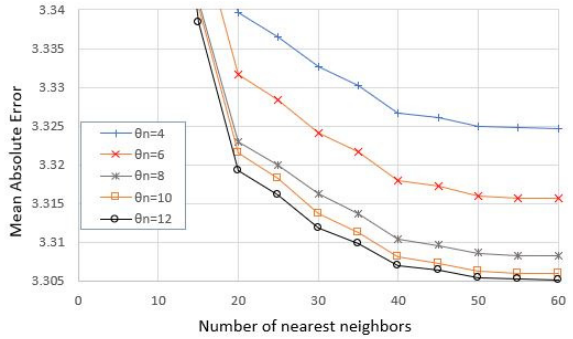
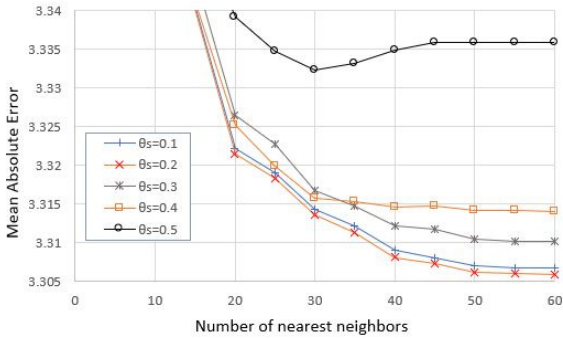


그림 2. 파라미터 θ_s (a)와 θ_n (b)의 변화에 따른 평균 절대 오차 성능
 Fig. 2. Mean absolute error with varying parameter values θ_s (a) and θ_n (b)

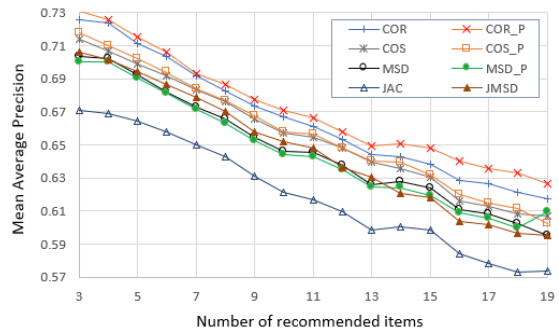
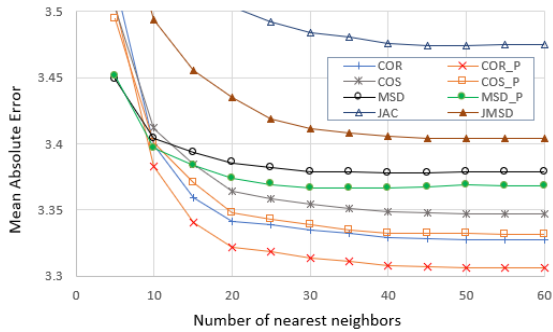


그림 3. 평균 절대 오차(MAE)와 평균 정밀도(MAP) 성능 비교
 Fig. 3. Performance comparison of mean absolute error(MAE) and mean average precision(MAP)

낮은 커버리지를 산출하게 된다. 따라서, 본 실험에서는 $\theta_n=10$, $\theta_s=0.2$ 로 하여 이후 실험을 진행하였다.

나. 성능 결과 비교

그림 3은 실험 방법들의 예측 정확도와 정밀도를 나타낸다. 인접 이웃 수가 증가할수록 미평가 항목에 대한 예측 오류는 점차 감소한 후 안정되는 것을 볼 수 있다. 같은 유사도 산출 방법을 사용한 경우, 제안 방법을 활용하여 산출된 예측치는 대응 방법에 비해 크게 성능이 향상되었는데, 이는 COR, COS, MSD 모두 해당되었다. 또한, JAC와 JMSD의 성능을 이들 방법들이 크게 능가함을 알 수 있다.

시스템에서 추천한 항목 리스트에 대한 적합도는 MAP의 정의에 따라, 추천 항목수가 많아지면서 감소한다. 그 양상은 MAE 결과와 거의 유사하게 나타났으나, 방법들 간의 성능 차이는 줄었음을 알 수 있다. 결과적으로, 제안한 평가치 예측 방법은 두 성능 평가 척도에서 모두 기존보다 우수한 결과를 보였다.

V. 결론

본 논문은 추천 시스템을 위한 이웃 기반의 협력 필터링 알고리즘에서 사용자가 미평가한 항목에 대한 평가치의 예측 방안을 제시하였다. 인접 이웃 수가 부족하여 예측치의 정확도가 저하되는 시스템의 본질적인 문제를 해결하기 위하여, 사용자 기반과 항목 기반의 예측치 산출 방법을 단계적으로 적용하였다. 제안 방법의 성능 실험을 통해, 전통적 유사도 척도 기반의 시스템들과 기존의 데이터 희소성 해결 방안들의 예측 성능과 정밀도 성능을 크게 향상시키는 결과를 보였다.

References

- [1] In-Kyu Park, Gyo-Seok Choi, "Regularized Optimization of Collaborative Filtering for Recommender System based on Big Data", The Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 21, No. 1, pp. 87-92, Feb.

2021. DOI: <https://doi.org/10.7236/JIIBC.2021.21.1.87>
- [2] Eun-Hee Jeong, Byung-Kwan Lee, "A Design of Customized Market Analysis Scheme Using SVM and Collaboration Filtering Scheme", The Journal of Korea Institute of Information, Electronics, and Communication Technology, Vol. 9, No. 6, pp. 609-616, Dec. 2016.
DOI: <http://dx.doi.org/10.17661/jkiict.2016.9.6.609>
- [3] Y.-J. Ahn, H.-Y. Cho, S.-J. Kang, "Customized Recipe Recommendation System Implemented in the form of a Chatbot," Journal of the Korea Academia-Industrial Cooperation Society, Vol. 21, No. 5, pp. 543-550, 2020.
DOI: <https://doi.org/10.5762/KAIS.2020.21.5.543>
- [4] E.-Y. Bae, S.-J. Yu, "A Study on the Cross Domain Recommendation System Using Adaptive Source Domain Selection," The Journal of Korean Institute of Information Technology, Vol. 17, No. 10, pp. 9-16, 2019. DOI : 10.14801/jkiit.2019.17.10.9
- [5] Z. Batmaz, A. Yurekli, A. Bilge, C. Kaleli, "A Review on Deep Learning for Recommender Systems: Challenges and Remedies", Artificial Intelligence Review, Vol. 52, No. 1, pp. 1-37, 2019.
DOI: <https://doi.org/10.1007/s10462-018-9654-y>
- [6] R. A. Mancisidor, M. Kampffmeyer, K. Aas, R. Jensen, "Learning Latent Representations of Bank Customers with the Variational Autoencoder", Expert Systems with Applications, Vol. 164, 2021.
DOI: <https://doi.org/10.1016/j.eswa.2020.114020>
- [7] M. Aamir, M. Bhusry, "Recommendation System: State of the Art Approach", International Journal Computer Applications, Vol. 120, No. 12, pp. 25-32, 2015.
DOI: <https://doi.org/10.5120/21281-4200>
- [8] M. Jalili, S. Ahmadian, M. Izadi, P. Moradi, M. Salehi, "Evaluating Collaborative Filtering Recommender Algorithms: A Survey", IEEE Access, Vol. 6, pp. 74003-74024, 2018.
DOI: <https://doi.org/10.1109/ACCESS.2018.2883742>
- [9] K. G. Saranya, G. S. Sadasivam, M. Chandralekha, "Performance Comparison of Different Similarity Measures for Collaborative Filtering Technique," Indian Journal of Science and Technology, Vol. 9, No. 29, Aug. 2016.
DOI: <https://doi.org/10.17485/ijst/2016/v9i29/91060>
- [10] S. Bag, S. K. Kumar, M. K. Tiwari, "An Efficient Recommendation Generation using Relevant Jaccard Similarity," Information Sciences, Vol. 483, pp. 53-64, 2019. DOI: <https://doi.org/10.1016/j.ins.2019.01.023>
- [11] S.-B. Sun, Z.-H. Zhang, X.-L. Dong, H.-R. Zhang, T.-J. Li, L. Zhang, F. Min, "Integrating Triangle and Jaccard Similarities for Recommendation", PLoS ONE, Vol. 12, No. 8, 2017.
DOI: <https://doi.org/10.1371/journal.pone.0183570>
- [12] S. Kosub, "A Note on the Triangle Inequality for the Jaccard Distance," Pattern Recognition Letters, Vol. 120, pp. 36-38, 2019.
DOI: <https://doi.org/10.1016/j.patrec.2018.12.007>
- [13] J. Bobadilla, F. Serradilla, J. Bernal, "A New Collaborative Filtering Metric that Improves the Behavior of Recommender Systems", Knowledge-Based Systems, Vol. 23, No. 6, pp. 520-528, 2010.
DOI: <https://doi.org/10.1016/j.knosys.2010.03.009>
- [14] B. Zhu, R. Hurtado, J. Bobadilla, F. Ortega, "An Efficient Recommender System Method Based on the Numerical Relevances and the Non-Numerical Structures of the Ratings," IEEE Access, Vol. 6, pp. 49935-49954, 2018.
DOI: <https://doi.org/10.1109/ACCESS.2018.2868464>
- [15] A. Iftikhar, M. A. Ghazanfar, M. Ayub, Z. Mehmood, M. Maqsood, "An Improved Product Recommendation Method for Collaborative Filtering", IEEE Access, Vol. 8, pp. 123841-123857, 2020.
DOI: <https://doi.org/10.1109/ACCESS.2020.3005953>

저 자 소 개

이 수 정(정회원)



- 1998.9~: 경인교육대학교 컴퓨터교육과 교수
- 1994.9~1998.4: 삼성전자 통신개발연구소 선임연구원
- 1994.8: Texas A&M 대학교 컴퓨터과학과 박사
- 1990.12: Texas A&M 대학교 컴퓨터과학과 석사
- 1985.2: 이화여자대학교 수학교육과 학사