Information Sciences 483 (2019) 53–64

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# An efficient recommendation generation using relevant Jaccard similarity

Sujoy Bag, Sri Krishna Kumar, Manoj Kumar Tiwari*

*Department of Industrial and Systems Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721 302, West Bengal, India*

ARTICLE INFO

*Article history:*
Received 23 July 2018
Revised 5 January 2019
Accepted 9 January 2019
Available online 11 January 2019

*Keywords:*
Recommender system
Collaborative filtering
Accuracy
Relevant Jaccard similarity
Data mining
Machine learning

ABSTRACT

In the literature, various collaborative filtering approaches have been developed to perform an efficient recommendation on top of reducing the search cost of the customers. The recommender system methods are concentrated on improving the accuracy and to achieve that goal they focused on formulating complex similarity approaches and neglect the computation time in their model. Furthermore, in order to compute the similarity metric, most of traditional similarity measures have only considered co-rated items and overlooked the total rating vector of the user or item. However, considering only co-rated items to measure similarity metrics in recommender system is an insignificant approach to identifying appropriate nearest neighbors in relatively sparse datasets. Therefore, in this research, two new simple but effective similarity models have been developed by considering all rating vectors of users to classify relevant neighborhoods and generate recommendations in a lower computation time. Moreover, MovieLens, a well-known dataset used in recommender system domain, is involved here to validate the performance of the proposed model. It seems that the proposed relevant Jaccard similarity perform more accurately and effectively to generate well recommendation than other traditional similarity models.



## 1. Introduction

Information-overload arduous the people while searching for the satisfactory product online. Thus, a decision support system is required to deal with information overload problems [14,32]. The recommender system plays an important role in the e-commerce industry for assisting customers in product selection [17,44]. Apart from e-commerce industry, recommender system has been widely used in various online business areas of travel, online broadcasting, scientific articles, advertising, news, movie, and music [5,27,36] etc. Furthermore, the recommendation system is recently applied in several important sectors like agriculture, traffic management, labor market, and medical [10,22,23,40] etc. Mostly, this system helps customer overcome dilemma by suggesting desired items based on the past behavior, relationship with others user, and item similarity. Establishing such system, customers' feedback data in the form of rating metrics is an influential factor to obtain appropriate items for the user. There are two types of feedback data present in the digital platform (i.e. implicit feedback and explicit feedback). The implicit feedback of a user is generated from the user's time-stamp on viewing a particular product, a number of clicks made in a page, types of product preferred in the past etc. Conversely, in the explicit feedback method, the user is asked to rate items or services that the user has previously viewed or purchased. Based on users' provided ratings, the model tries to group similar users or alternative items, by building a user-user or item-item similarity

* Corresponding author.
*E-mail address:* mktiwari9@iem.iitkgp.ernet.in (M.K. Tiwari).

https://doi.org/10.1016/j.ins.2019.01.023
0020-0255/

index. Then the algorithm predicts a suitable list of items for recommending to the user. For example, Last.fm provides a listener with the song from a list of singers the user or similar users have listened repeatedly in the past. Similarly, MovieLens helps the user find personalized movies based on the users' custom taste profile which is generated from the user's rating matrix.

To handle the huge amount of rating data in this system, various studies have been carried out since last few decades. Mainly, three filtering techniques such as content-based, collaborative and hybrid filtering are presented in the recommender system literature to filter records and identify the relevant information. Content-based and collaborative filtering (CF) work based on user/item profile favored in the past and user/item similarity metrics. Hybrid Filtering is the combination of content-based and collaborative filtering. In this research, collaborative filtering is often chosen for further improvement as the upgraded version of this algorithm can be easily applied and re-implemented in various domains and many times perform better than the content-based and hybrid filtering. In general, collaborative filtering predicts the rating for an unrated item of a target user based on the rating patterns of rated items of $k$-nearest neighbors and ultimately suggests preferred items to the user.

However, most of the advanced collaborative filtering methods [3,37] become complex specifically while dealing with fewer number of co-rated items. Traditional similarity measures are simple and performed effectively through it has some inadequacies in identifying appropriate nearest neighbors. Thus in this research, two new simple but effective relevant Jaccard similarity and relevant Jaccard mean square distance have been developed to classify relevant neighborhoods and generate better recommendations in lower computation time. Moreover, the past similarity measures that rely on only co-rated items to form neighborhood and neglect the total rating vector of the user or item, have various limitations and poor performance in recommendation generation. Therefore, the proposed study has been utilized all the rating vectors to make better prediction and perform well in recommendation generation. The proposed similarity model has given priority in minimum number of un-co-rated items of the targeted user and maximum number of co-rated[1] and un-co-rated[2] items of nearest neighbors.

In this research, initially, the limitations of the existing traditional similarity methods which lead to inaccurate prediction have been presented with a proper explanation. Further, the Jaccard similarity approach has been derived to simplify and formulate drawbacks of this approach in the context of the recommendation system. In order to formulate new relevant Jaccard similarity model, an illustration with suitable examples of the Jaccard similarity method in recommender system framework has been presented in this study. Furthermore, the proposed relevant Jaccard similarity and mean square distance similarity have been merged to form another new similarity metric called relevant Jaccard mean square distance (RJMSD) similarity.

Proposed relevant Jaccard similarity model can be performed efficiently in many application and research domain as it is an improved version of the Jaccard similarity method which is a widely recognized traditional similarity metric. The obtained results show that newly formed relevant Jaccard similarity and RJMSD perform better than the existing traditional similarity metrics including Jaccard similarity and Jaccard mean square distance similarity.

Rest sections of this paper are organized as follows, in Section 2, details literature review including fundamental mathematical models of collaborative filtering are described. Further, several limitations of traditional similarity approaches are specified in Section 3. Additionally, two new similarity approaches are proposed in Section 4. Moreover, experimentations and results are shown in Section 5. Lastly, the article is concluded after discussing the study and mentioning limitations of the work with enabling future research scope in Section 6.

## 2. Literature review

Content-based recommendation and collaborative filtering are mainly two widely used techniques in recommendation generation. Content-based method mostly takes into account the implicit rating by text mining process and makes a recommendation, whereas collaborative filtering considers only explicit ratings of users [15,18,28]. Furthermore, there are two types of collaborative filtering techniques frequently used in the recommendation system domain such as model-based and memory-based collaborative filtering. Model-based method develops a user model utilizing ratings of each user to evaluate the expected value of unrated items [29,42]. On the other hand, memory-based method utilizes similarity measure computed from the explicit user rating to identify neighborhoods and perform prediction [7,46].

However, collaborative filtering predicts random rating values because of the stochastic prediction methodology, dynamic rating data, and subjectivity users. Thus, an accurate rating prediction is an insignificant methodology in the recommender system domain. Still, researchers are constantly attempted to improve the performance of collaborative filtering in terms of accuracy metrics. In this regard, Ren and Wang (2018) applied a support vector machine (SVM) in the collaborative filtering (CF) method to increase the quality and efficiency of the web service recommendation system [39]. Furthermore, Liu and Wu (2016) have proposed a scoring function and latent collaborative relations to provide a high-speed recommendation system while maintaining accuracy and coverage [30]. Similarly, Polatidis and Georgiadis (2017) introduced a dynamic multi-level CF approach based on positive and negative adjustments to enhance the quality of the user experience [38]. Recently, a hybrid

[1] Co-rated means the number of ratings collaboratively given by any two users.

[2] Un-co-rated ratings = [total ratings which are given by a specific user - number of co-rated ratings].

user similarity model has been developed with integrating the Kullback–Leibler (KL) divergence and adjusted Proximity–Significance–Singularity model [48]. Moreover, Tran et al. (2016) proposed a formal probabilistic framework to estimate the correlation between user-user and item-item by hosting a sparsity-inducing algorithm in Markov random fields (MRF) [47]. Additionally, Pan and Chen (2016) included rich interactions in the group Bayesian personalized ranking and adopted a set of items to generate better-quality recommendations [33,34].

Further, one-class and matrix factorization-based algorithms are presented in many recommender system articles [9]. For instance, Li and Ou (2016) developed a pairwise probabilistic matrix factorization with the help of RankRLS to learn the relative preference for items from the users' implicit feedback [25]. Besides, various hybrid approaches with the help of neural networks [31,35,41] are presented in the literature to improve the performance of the decision support system. Shi et al. (2017) considered recommender systems as a bipartite network and adopt a network evolution method to evaluate the long-term performance of various CF-based recommender systems [44]. Collaborative filtering has also been performed over social network and users' profiles to conduct a group recommendation system [8]. Li et al. (2017) introduced a hybrid collaborative filtering approach to predict the social influences of users on upcoming occasions in event-based social networks [24]. In addition, this filtering method has been implemented in many social network applications. Shahmohammadi et al. (2016) have developed a new collaborative link prediction technique for activity prediction and recommendation on Facebook [43]. Largely, social network mining, sentiment analysis and search modeling [4,11,21] are an important methodology to analysis the consumer behavior and recommend personalized products.

In the literature, various traditional similarity approaches like cosine similarity, Pearson's correlation, Jaccard similarity etc. and several data mining techniques are presented to classify nearest neighbors and generate recommendations. Although, most of the techniques are complex or time-consuming and also further improvements have been observed. The purpose of the proposed similarity model is to identify appropriate nearest neighbors of the targeted user and predict accurate ratings of unrated items.

### *2.1. Fundamental mathematical models of collaborative filtering*

Mathematical formulas are presented in this section, which are performed in the proposed model. Primarily, neighborhood-based prediction technique has been described and then various traditional similarity measures are presented and shown their limitations in recommendations generation.

#### *2.1.1. Neighborhood-based prediction technique*

Let us consider $U = \{u_1, u_2, u_3, .....u_m\}$ as a user set, $I = \{i_1, i_2, i_3, ......i_n\}$ as an item set and $R = \{R(1, 1), R(1, 2), ......R(m, n)\}$ as a set of user-item rating pairs where, $R(u,\ i)$ means rating of the user $u$ for item $i$. Generally, $R(u, i) \in W = \{w_1, w_2, w_3, ......, w_N\}$ set of discrete rating scores where a higher rating indicates the user strongly likes that item. Similarity indices between users or items are calculated from the user-item rating matrix to group similar users and items for appropriate recommendations. The fundamental role of a user-user similarity index is to classify users who have given similar ratings to a set of items in the past. Likewise, an item-item similarity index is calculated by taking into account the set of items that are co-rated by a set of users. The prediction is then generated for all unrated items by forming similarity indices of targeted users with $k$-nearest neighbors. The unknown rating is predicted based on user-user similarity as,

$$R^*(u, i) = \overline{R(u)} + \frac{\sum_{v \in S(u)} Sim(u, v) \cdot \left(R(v, i) - \overline{R(v)}\right)}{\sum_{v \in S(u)} |Sim(u, v)|} \tag{1.1}$$

Where, $R^*(u,\ i)$ is the predicted rating of the targeted user $u$ for item $i$ and $\overline{R(u)}$ is the mean rating of user $u$. $\underline{S(u)}$ is the number of top similar users who have also rated item $i$. $R(v,\ i)$ is rating of the nearest neighbor $v$ for item $i$. $\overline{R(v)}$ is the mean rating of the nearest neighbor $v$. $Sim(u,\ v)$ is the similarity index between the user $u$ and its nearest neighbor $v$. The absolute sign for similarity value is used in the denominator to avoid the negative correlation between the targeted user and nearest neighbors. The motivation behind $(R(v, i) - \overline{R(v)})$ is that if the neighbor $v$ has rated the item above average, then it will add to the average rating of the user $u$ . Similarly, the logic behind multiplying $Sim(u,\ v)$ with $(R(v, i) - \overline{R(v)})$ is that if the similarity between the user and its neighbor $v$ is very high, then the user's rating prediction will be heavily influenced by the neighbor $v$ and vice versa.

#### *2.1.2. The traditional similarity measures*

Similarity measure in the recommender system is the statistical measure of how two users and items are related to each other. There are several traditional similarity metrics such as Cosine (COS), Pearson's Correlation (COR), Constrained Pearson's Correlation (CPC), Mean Squared Difference (MSD), Jaccard, JMSD etc. [37].

*2.1.2.1. Cosine similarity (COS).* Cosine similarity measures the angle between two rated vectors where the smaller angle indicates greater similarity and higher angle show lesser similarity [3].

$$Sim(u,v)^{COS} = \frac{\sum\limits_{i\in I(u,v)} R(u,i)\cdot R(v,i)}{\sqrt{\sum\limits_{i\in I(u,v)} R(u,i)^2}\cdot\sqrt{\sum\limits_{i\in I(u,v)} R(v,i)^2}} \tag{1.2}$$

Where $R(u, i)$ is the rating of the item $i$ given by user $u$ and $I(u, v)$ is the number of co-rated items of users $u$ and $v$. The range of cosine similarity is 0 to 1, where higher value signifies the closest similarity between users $u$ and $v$.

*2.1.2.2. Pearson's correlation (COR).* The estimated ratio of the cross product of overrating or underrating of means divided by the product of the sum of squares of mean rating difference is called the Pearson's correlation coefficient [3].

$$Sim(u,v)^{COR} = \frac{\sum\limits_{i\in I(u,v)} \left(R(u,i)-\overline{R(u)}\right)\cdot\left(R(v,i)-\overline{R(v)}\right)}{\sqrt{\sum\limits_{i\in I(u,v)} \left(R(u,i)-\overline{R(u)}\right)^2}\cdot\sqrt{\sum\limits_{i\in I(u,v)} \left(R(v,i)-\overline{R(v)}\right)^2}} \tag{1.3}$$

Where $R(u, i)$ is the rating of item $i$ given by the user $u$ and $I(u, v)$ is the number of co-rated items of users $u$ and $v$. $\overline{R(u)}$ is the average rating of the user $u$. A value of $-1$ in the Pearson's similarity indicates a negative correlation between users, 0 indicates the neutral correlation between users and $+1$ indicates a positive correlation between users.

*2.1.2.3. Constrained pearson's correlation (CPC).* Constrained Pearson's Correlation is the modified version of the Pearson's Correlation [3]. This similarity allows only pairs of ratings on the same positive or negative side, to increase in the correlation.

$$Sim(u,v)^{CPC} = \frac{\sum\limits_{i\in I(u,v)} (R(u,i)-R_m)\cdot(R(v,i)-R_m)}{\sqrt{\sum\limits_{i\in I(u,v)} (R(u,i)-R_m)^2}\cdot\sqrt{\sum\limits_{i\in I(u,v)} (R(v,i)-R_m)^2}} \tag{1.4}$$

Where $R(u, i)$ and $R(v, i)$ are the rating of the item $i$ given by the user $u$ and $v$ respectively. $R_m$ is the median value in the rating scale (e.g. three on the rating scale of 5). $I(u,v)$ denotes the co-rated items of users $u$ and $v$.

*2.1.2.4. Jaccard similarity.* Jaccard similarity index mainly focuses on global ratings. It is the ratio of the proportion of the cardinality of co-rated items to the cardinality of all items rated by both the user.

$$Sim(u,v)^{Jaccard} = \frac{(I_u\cap I_v)}{(I_u\cup I_v)} \tag{1.5}$$

Where $I_u$ and $I_v$ are the set of items rated by users $u$ and $v$ respectively.

*2.1.2.5. Mean square distance (MSD).* Similarly, Mean Square Distance (MSD) between two users is calculated by the ratio of sum square of the difference of ratings on co-rated items and the cardinality of co-rated items. The Mean square Similarity is then calculated by subtracting MSD from 1.

$$Sim(u,v)^{MSD} = 1-\frac{\sum_{i\in I(u,v)} (R(u,i)-R(v,i))^2}{|I(u,v)|} \tag{1.6}$$

Where $R(u, i)$ and $R(v, i)$ are the rating of the item $i$ given by the user $u$ and $v$ respectively. $I(u,v)$ indicates the co-rated items of users $u$ and $v$.

*2.1.2.6. Jaccard mean square distance (JMSD).* Jaccard Mean Square Distance measures the partial similarity of Jaccard and MSD, which is generated from the multiplication of these two similarity measures.

$$\begin{aligned} Sim(u,v)^{JMSD} &= \left(Sim(u,v)^{Jaccard}\right)\cdot\left(Sim(u,v)^{MSD}\right)\\ &= \left(\tfrac{(I_u\cap I_v)}{(I_u\cup I_v)}\right)\cdot\left(1-\tfrac{\sum_{i\in I(u,v)} (R(u,i)-R(v,i))^2}{|I(u,v)|}\right) \end{aligned} \tag{1.7}$$

*2.1.2.7. Weighted regularized matrix factorization (WRMF).* This WRMF method extracts confidence levels and preferences from the implicit feedback by using the confidence value as a weight [16,20].

*2.1.2.8. Factored item similarity models (FISM).* FISM approach is an item-based method which learns to rank the item-item similarity matrix as the product of two low dimensional latent factor matrices [19,20].

## 3. Limitations of traditional similarity measures

Though different similarity measures are adopted by different online companies for their recommendation system, still a lot of loopholes have been noticed in various situations, which leads inaccurate prediction.

(1) Let $U_1 = (2, 0, 3, 0)$ and $U_2 = (5, 2, 0, 2)$ are the rating vectors of two users where only one co-rated item presents. It is noticed that the Pearson's correlation coefficient cannot be determined as because denominator becomes zero. Likewise, cosine similarity yields 100% similarity regardless of an actual matching.
(2) Further, let $U_1 = (2, 1, 3, 2)$ and $U_2 = (1, 2, 2, 3)$ be the rating array of two users. Even though both the users are highly similar, the Pearson's correlation coefficient generates zero similarity indexes.
(3) In another situation, let $U_1 = (2, 2, 0, 1)$ and $U_2 = (4, 4, 0, 2)$ be ratings of two users. In this case, the cosine similarity index computes a similarity value of 1 which is an insignificant similarity valuation. Cosine similarity always yields a very high similarity (i.e. 1) when ratings are multiple of each other because in that scenario geometrically they overlap each other on the same straight line.
(4) Although, Jaccard similarity takes into account the global ratings still in the absence of local co-rated items it becomes zero. Additionally, Jaccard similarity does not consider the absolute rating value in similarity computation time. Let, consider $U_1 = (5, 5, 4, 3)$ and $U_2 = (1, 2, 2, 1)$, here Jaccard similarity index computes a totally contrary similarity index of 1, where real similarity is quite low. Furthermore, ignoring the total rating length items and focusing on only co-rated items yields loss of information.

## 4. Proposed similarity methods

Jaccard similarity, simplicity in nature and user-friendly in implementation, is applied in various domain and performed adequately. The Jaccard coefficient measures the similarity between the finite sample sets. Jaccard similarity is the ratio between the size of the intersection and the size of the union of sample sets. Set representation of rated items $I_u$ and $I_v$ by users $u$ and $v$ is presented in Fig. 1. Jaccard only counts the number of common ratings between two users to evaluate the similarity metric [13,37].

$$sim(u, v)^{Jaccard} = \frac{|I_u \cap I_v|}{|I_u \cup I_v|} \tag{1.8}$$

Where $I_u$ and $I_v$ are the set of items rated by users $u$ and $v$ respectively.

As, $I_u$ and $I_v$ are not mutually exclusive, hence addition rule theorem is applied here to form, $|I_u \cup I_v| = |I_u| + |I_v| - |I_u \cap I_v|$ Whereas, $|I_u|$ and $|I_v|$ are the cardinality of the sets $I_u$ and $I_v$, respectively.

$$\text{Therefore now, } sim(u, v)^{Jaccard} = \frac{|I_u \cap I_v|}{|I_u| + |I_v| - |I_u \cap I_v|} \tag{1.9}$$

Suppose, $|\overline{I_u}|$ and $|\overline{I_v}|$ are the cardinality of the set of items un-co-rated by users $u$ and $v$ respectively. Hence, $|\overline{I_u}| = |I_u| - |I_u \cap I_v|$ and $|\overline{I_v}| = |I_v| - |I_u \cap I_v|$.

As a result, the Jaccard similarity can be written as,

$$sim(u, v)^{Jaccard} = \frac{|I_u \cap I_v|}{\left(\left|\overline{I_u}\right| + |I_u \cap I_v|\right) + \left(\left|\overline{I_v}\right| + |I_u \cap I_v|\right) - |I_u \cap I_v|} = \frac{|I_u \cap I_v|}{\left|\overline{I_u}\right| + \left|\overline{I_v}\right| + |I_u \cap I_v|} \tag{1.10}$$

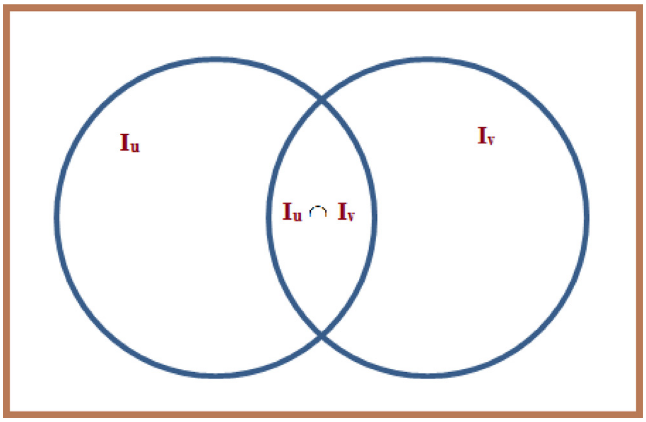


**Fig. 1.** Set representation of rated items $I_u$ and $I_v$ by users $u$ and $v$ respectively.

**Table 1**
Number of un-co-rated items of different nearest neighbors for the targeted user $U_1$.

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ | # of un-co-rated items of different nearest neighbors for the targeted user $U_1$ |
|---|---|---|---|---|---|---|---|---|---|
| $U_1$ | 3 | ? | 4 | 2 | ? | 5 | 1 | 1 | N/A |
| $U_2$ | 3 | - | - | 2 | - | - | - | 1 | 0 |
| $U_3$ | 3 | - | 4 | - | - | 5 | - | - | 0 |
| $U_4$ | - | - | - | 2 | - | - | 1 | 1 | 0 |
| $U_5$ | 4 | 2 | - | 2 | 5 | 5 | - | 2 | 2 |
| $U_6$ | 4 | 3 | - | - | 4 | - | 1 | 2 | 2 |

**Table 2**
Number of un-co-rated items of the targeted user $U_1$ for different nearest neighbors.

| | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ | $I_7$ | $I_8$ | # of un-co-rated items of the targeted user $U_1$ for different nearest neighbors |
|---|---|---|---|---|---|---|---|---|---|
| $U_1$ | 3 | - | 2 | - | 5 | 5 | 4 | ? | N/A |
| $U_2$ | 2 | 1 | - | 5 | - | - | - | 1 | 4 |
| $U_3$ | - | 2 | 1 | - | - | - | - | 2 | 4 |
| $U_4$ | - | - | - | 4 | - | 4 | - | 1 | 4 |
| $U_5$ | 3 | 2 | 3 | - | 4 | 5 | 3 | 4 | 0 |
| $U_6$ | 4 | - | 2 | - | - | 5 | 3 | 5 | 1 |

Divide both the numerator and denominator by $|I_u \cap I_v|$, thus now Jaccard similarity,

$$sim(u, v)^{Jaccard} = \frac{1}{\frac{|\overline{I_u}|}{|I_u \cap I_v|} + \frac{|\overline{I_v}|}{|I_u \cap I_v|} + 1} = \frac{1}{1 + \frac{|\overline{I_u}| + |\overline{I_v}|}{|I_u \cap I_v|}} \tag{1.11}$$

This equation express, $sim(u,v)^{Jaccard} \propto |I_u \cap I_v|$ and $sim(u, v)^{Jaccard} \propto (|\overline{I_u}| + |\overline{I_v}|)^{-1}$. However, sometimes this demonstration specifically $sim(u, v)^{Jaccard} \propto (|\overline{I_u}| + |\overline{I_v}|)^{-1}$ performs an inappropriate role in the Jaccard similarity model for suitable recommendations generation to the customer. A certain amount of inadequacies and further improvements have been noticed in this similarity model while applied to measure similarity metric using minimum number of co-rated items.

*4.1. Illustration with examples to formulate new similarity model*

Traditional recommender systems consider only co-rated items to determine similarity to identify nearest neighbors. Collaborating filtering predicts the rating of unrated items using ratings of the same item given by nearest neighbors, but sometimes this procedure is unsuitable to predict ratings of unrated items. Suppose any user is highly similar but rated only co-rated items. In that case, the similar user is useless to predict rate.

In Tables 1 and 2, two samples of the user-item rating matrix are presented to illustrate limitations of Jaccard similarity where the missing rating represents the symbol "-". Table 1 shows the number of un-co-rated items of different nearest neighbors for the targeted user $U_1$. In this example, the targeted user $U_1$ is highly similar to the users $U_2$, $U_3$, and $U_4$. However, these nearest neighbors are useless to predict rates of unrated items $I_2$ and $I_5$ of the user $U_1$ as the number of un-co-rated items of nearest neighbors $U_2$, $U_3$, and $U_4$ are 0. On the other hand, users $U_5$ and $U_6$ are comparatively less similar to the user $U_1$. Although these two users are effective to predict rates of unrated items of the user $U_1$ as users $U_5$ and $U_6$ have already given the ratings on items $I_2$ and $I_5$. This situation indicates that the number of un-co-rated items of nearest neighbors is proportional to generate an informative similarity metric (i.e. $sim(u, v)^{Jaccard} \propto (|\overline{I_v}|)$).

**Table 3**
Summary of the MovieLens Dataset.

| Profile | 943 users 1682 movies 100,000 ratings |
|---|---|
| Ratings per user | On more than 100 movies per user |
| Description | Ratings of movies in the scale of 1–5 |

Another situation is presented in Table 2 where numbers of un-co-rated items of the targeted user $U_1$ for different nearest neighbors have been shown. In this scenario, only one item rating provided by the user $U_1$ is used to compute similarity with users $U_2$, $U_3$, and $U_4$. Total 4 numbers of un-co-rated items of targeted users $U_1$ are presented for all nearest neighbors $U_2$, $U_3$, and $U_4$. On the other hand, a number of un-co-rated items of targeted users $U_1$ for nearest neighbors $U_5$ and $U_6$ are 0 and 1, respectively. Here, all items rating given by the user $U_1$ is informative to generate adequate similarity metric between users $U_5$ and $U_6$. This condition implies that the number of un-co-rated items of targeted users is inversely proportional to generate an informative similarity metric (i.e. $sim(u,\ v)^{Jaccard} \propto \ (|\overline{I_u}|)^{-1}$).

Another situation which Jaccard similarity generally consider is that if a number of co-rated items increase then the significance of similarity value also increases proportionally (i.e. $sim(u,v)^{Jaccard} \propto |I_u \cap I_v|$).

Thus, only similarity is insufficient to find the effective nearest neighbor and predict unrated items rating. As a result, traditional nearest neighborhood-based collaborative filtering used an extreme number of nearest neighbors to normalize the effect. Although, recommender system turn into complicated, erroneous, and time-consuming while it used an excessive number of nearest neighbors. Therefore, a new similarity model called relevant Jaccard similarity has been formed based on generated conditions from the illustration. Now, relevant Jaccard similarity can be measured from the following equation which follow all generated conditions (i.e. $sim(u,\ v)^{Jaccard} \propto \ (|\overline{I_v}|)$, $sim(u,\ v)^{Jaccard} \propto \ (|\overline{I_u}|)^{-1}$, and $sim(u,v)^{Jaccard} \propto |I_u \cap I_v|$).

$$sim(u,\ v)^{RJaccard} = \frac{1}{1 + \frac{1}{|I_u \cap I_v|} + \frac{|\overline{I_u}|}{1+|\overline{I_u}|} + \frac{1}{1+|\overline{I_v}|}} \tag{1.12}$$

Where, $if\ |I_u \cap I_v| = 0\ then\ sim(u,\ v)^{RJaccard} = 0$

*4.2. New similarity models*

In this research, two new similarity models have been designed, and furthermore, the convenient similarity model can be formed by utilizing the proposed similarity model.

*4.2.1. Relevant Jaccard similarity (RJaccard)*

It is an improved version of the Jaccard Similarity model.

$$Sim(u,v)^{RJaccard} = \frac{1}{1 + \frac{1}{|I_u \cap I_v|} + \frac{|\overline{I_u}|}{1+|\overline{I_u}|} + \frac{1}{1+|\overline{I_v}|}} \tag{1.13}$$

Where, $if\ |I_u \cap I_v| = 0\ then\ sim(u,\ v)^{RJaccard} = 0$

*4.2.2. Relevant Jaccard mean square distance (RJMSD)*

In this research, RJMSD generated by multiplying the two similarity metrics of RJaccard and MSD.

$$\begin{aligned} Sim(u,v)^{RJMSD} &= Sim(u,v)^{RJaccard} \times Sim(u,v)^{MSD} \\ &= \left( \frac{1}{1+\frac{1}{|I_u \cap I_v|} + \frac{|\overline{I_u}|}{1+|\overline{I_u}|} + \frac{1}{1+|\overline{I_v}|}} \right) \times \left( 1 - \frac{\sum_{i \in I(u,v)} (R(u,i) - R(v,i))^2}{|I(u,v)|} \right) \end{aligned} \tag{1.14}$$

Where, $if\ |I_u \cap I_v| = 0\ then\ sim(u,\ v)^{RJMSD} = 0$

## 5. Experimentation and results

Initially, experimental setups including a description of the used dataset, efficiency measurement metrics have been explained and then the performance of the proposed model is shown in the rest of this section.

*5.1. Dataset*

In this research, the MovieLens dataset [12] has been carried out to evaluate the performance of the prediction model, the detailed of the dataset is shown in Table 3. The dataset is separated into two parts as 70% and 30% for training and testing the model respectively.

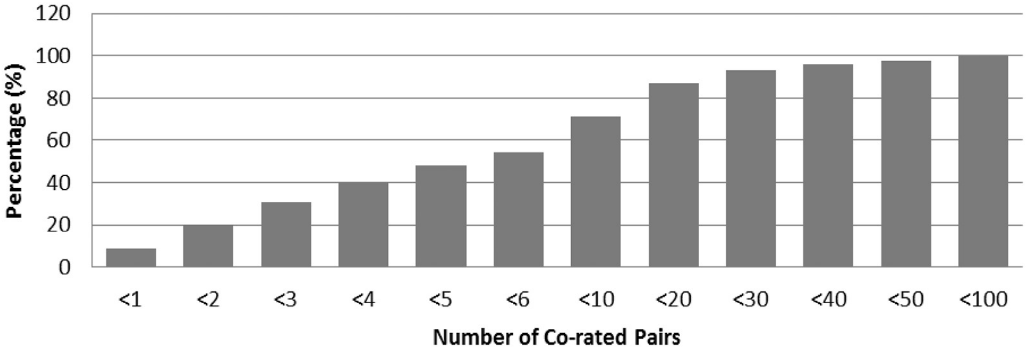


**Fig. 2.** Insights of the Training dataset.

**Algorithm 1.**
Rating prediction using relevant Jaccard similarity model.

***Input***: *User - itemrating metric*
***Output***: *Value of predicted ratings*
$sim\,(u,\,v) = [\;],\ R^*(u,i) = [\;]$
***for*** *the user* $u = 1$ *to* $U$ ***do***
    ***for*** *the user* $v = 1$ *to* $U$ ***do***
        ***if*** $u! = v$ *and* $|I_u \cap I_v| > 0$ ***then***
            $sim(u,v) = \frac{1}{1 + \frac{1}{|I_u \cap I_v|} + \frac{|\overline{I_u}|}{1+|\overline{I_u}|} + \frac{1}{1+|\overline{I_v}|}}$
        ***else***
            $sim(u,v) = 0$
        ***end if***
        ***end for***
    ***end for***
    ***for*** *the user* $u = 1$ *to* $U$ ***do***
        ***for*** *the item* $i = 1$ *to* $I$ ***do***
            ***if*** $R\,(u,\,i) == 0$ *and* $S\,(u) > 0$ ***then***
                $R^*(u,i) = \overline{R(u)} + \frac{\sum_{v \in S(u)} Sim(u,v)\cdot(R(v,i) - \overline{R(u)})}{\sum_{v \in S(u)} |Sim(u,v)|}$
        ***else***
            $R^*(u,i) = 0$
        ***end if***
        ***end for***
***end for***

Insights of the training dataset are shown in Fig. 2 where the number of co-rated pairs exists in the dataset is presented. A total number of user pairs of this dataset are 444,153. 8.87% user pairs have no co-rated items, and share of having 1, 2, 3, 4, 5 numbers of user pairs are quite high. More than 50% of couples have less than equals to 5 co-rated items. 71.43% of couples have less than 10 co-rated items. We have observed a decreasing trend in the number of user pairs while increasing the number of co-rated items. There is only 0.1% of the pair have more than or equal to 100 co-rated items, whereas the share of pairs having more than or equal to 150 co-rated items is negligible. Thus, rating prediction for unrated items is challenging from this relatively sparse dataset.

### *5.2. Prediction algorithm*

In this section, two rating prediction algorithms are presented in Algorithm 1 and 2, which help the recommender system in identifying desired products for the customer. In Algorithm 1, relevant Jaccard similarity model is performed to compute an efficient similarity metric and recommendations. Subsequently, in Algorithm 2, relevant Jaccard mean square distance (RJMSD) is implicated to compute harmonize of local and global similarity metrics between two users. Further, RJMSD is used in prediction of unrated items of each user to provide appropriate item recommendations.

### *5.3. Evaluation metrics*

Evaluation metrics are used to assess the performance of any machine learning algorithm [26]. In this study, F1-Measure and aggregate diversity are considered as an evaluation metric to validate the performance of the proposed model. Measuring the precision and the recall are required before calculating the F1-Measure as it is the harmonic mean of precision and recall. Precision is the ratio of the number of items are actually relevant to the user to the total number of items are predicted relevant to the user. In the literature of relevant items are items that have got higher ratings. The recall is the

**Algorithm 2.**
Rating prediction using relevant Jaccard mean square distance similarity model.

***Input:*** *User - item rating metric*
***Output:*** *Value of predicted ratings*
$sim(u, v) = [\,],\ R^*(u, i) = [\,]$
***for*** *the user* $u = 1$ *to* $U$ ***do***
    ***for*** *the user* $v = 1$ *to* $U$ ***do***
        ***if*** $u! = v$ *and* $|I_u \cap I_v| > 0$ ***then***
            $sim(u,v) = \left( \frac{1}{1 + \frac{1}{|I_u \cap I_v|} + \frac{|\overline{I_u}|}{1 + |\overline{I_u}|} + \frac{1}{1 + |\overline{I_v}|}} \right) \times \left( 1 - \frac{\sum_{i \in I(u,v)} (R(u,i) - R(u,i))^2}{|I(u,v)|} \right)$
        ***else***
            $sim(u,v) = 0$
        ***end if***
        ***end for***
    ***end for***
    ***for*** *the user* $u = 1$ *to* $U$ ***do***
        ***for*** *the item* $i = 1$ *to* $I$ ***do***
            ***if*** $R(u, i) == 0$ *and* $S(u) > 0$ ***then***
                $R^*(u,i) = \overline{R(u)} + \frac{\sum_{v \in S(u)} Sim(u,v) \cdot (R(v,i) - \overline{R(u)})}{\sum_{v \in S(u)} |Sim(u,v)|}$
        ***else***
            $R^*(u,i) = 0$
        ***end if***
        ***end for***
***end for***

measure of the proportion of items that are predicted well to the total number of items that are actually preferable. The evaluation metrics such as precision, recall, and F1-measure are calculated [26,45] from the following equations,

$$\boldsymbol{Precision} = \frac{TP}{TP + FP} \tag{1.15}$$

$$\boldsymbol{Recall} = \frac{FN}{TP + FP} \tag{1.16}$$

$$\boldsymbol{F_1 - Measure} = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{1.17}$$

**Aggregate diversity:** The aggregate diversity is the number of distinct elements recommended to the entire user segment, which can be computed [1,2] from the equation,

$$diversity = \left| \bigcup\nolimits_{u \in U} L_n(u) \right| \tag{1.18}$$

Where $u$ is any particular user, $U$ is the total user in the dataset and $L_n(u)$ is the list of relevant items recommended to the user of $u$.

### 5.4. Performance of the proposed model

Measuring the performance of any prediction model is an important aspect of machine learning domain which is performed here by computing F1-Measure and aggregate diversity of the recommender system. All traditional similarity methods including proposed two approaches have classified that the number of recommended items decrease while the threshold value increase, and vice versa. On the other hand, F1-Measure is proportional to the number of recommended items. The threshold value is the acceptable predicted rating which can be in any range of rating vector (here, 1–5). Higher threshold value indicates that the system recommends most likely items to the customer. In this study, variations of F1 measure and aggregate diversity are observed in threshold values of 3.5–4.5 whereas four instances of nearest neighbors such as 5, 20, 50, and 100 are considered to draw variants performance of the model. F1-Measures of WRMF, FISM and various traditional similarities (i.e. Jaccard, MSD, JMSD, COS, COR, and CPC) have been presented in Fig. 3 to compare the performance of proposed similarity metrics. Likewise, aggregate diversities of traditional similarity metrics have also been presented in Fig. 4 to compare the obtained results.

It is noticed that proposed RJaccard and RJMSD approaches perform remarkably in terms of F1-Measure and aggregate diversity when considering 5 nearest neighbors. Furthermore, it has been observed that if moderate likely items are recommended to customers in the system (i.e. threshold value 3.5) then all similarity models perform approximately equally. However, if the system wishes to recommend most likely items to customers then it is better to use proposed RJaccard or RJMSD similarity methods.

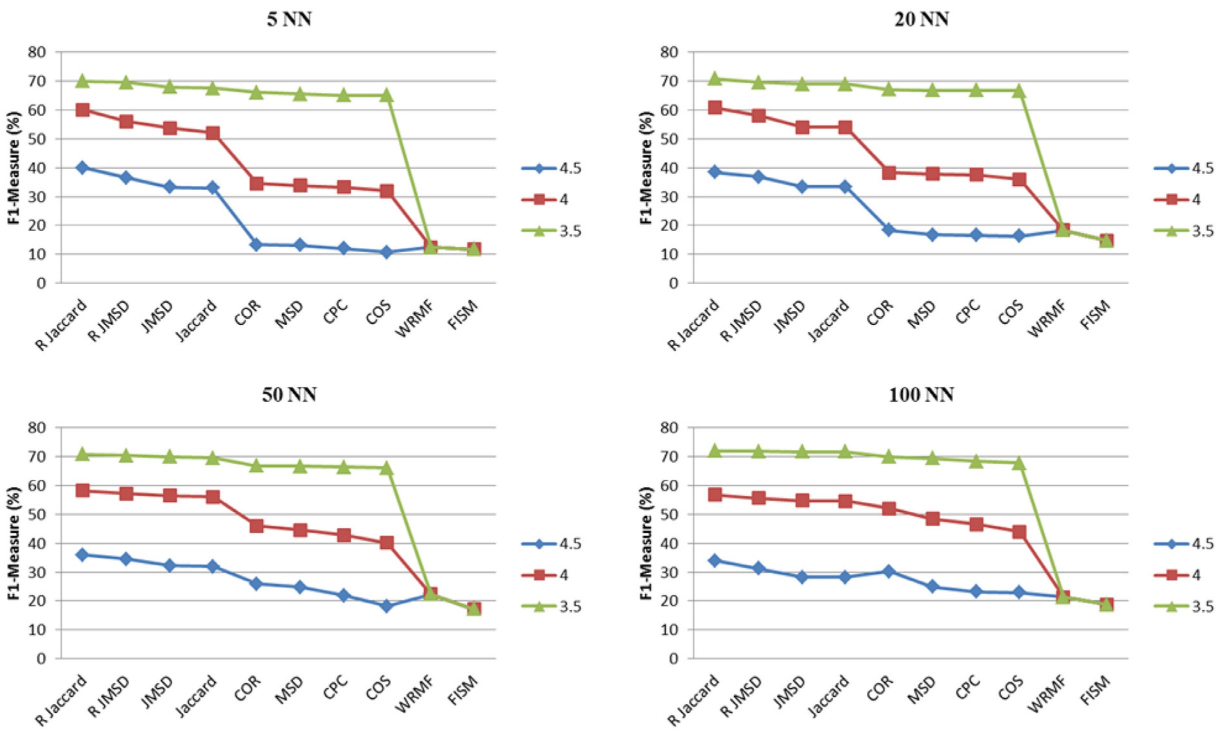


Fig. 3. F1 measures of proposed and relevant approaches.

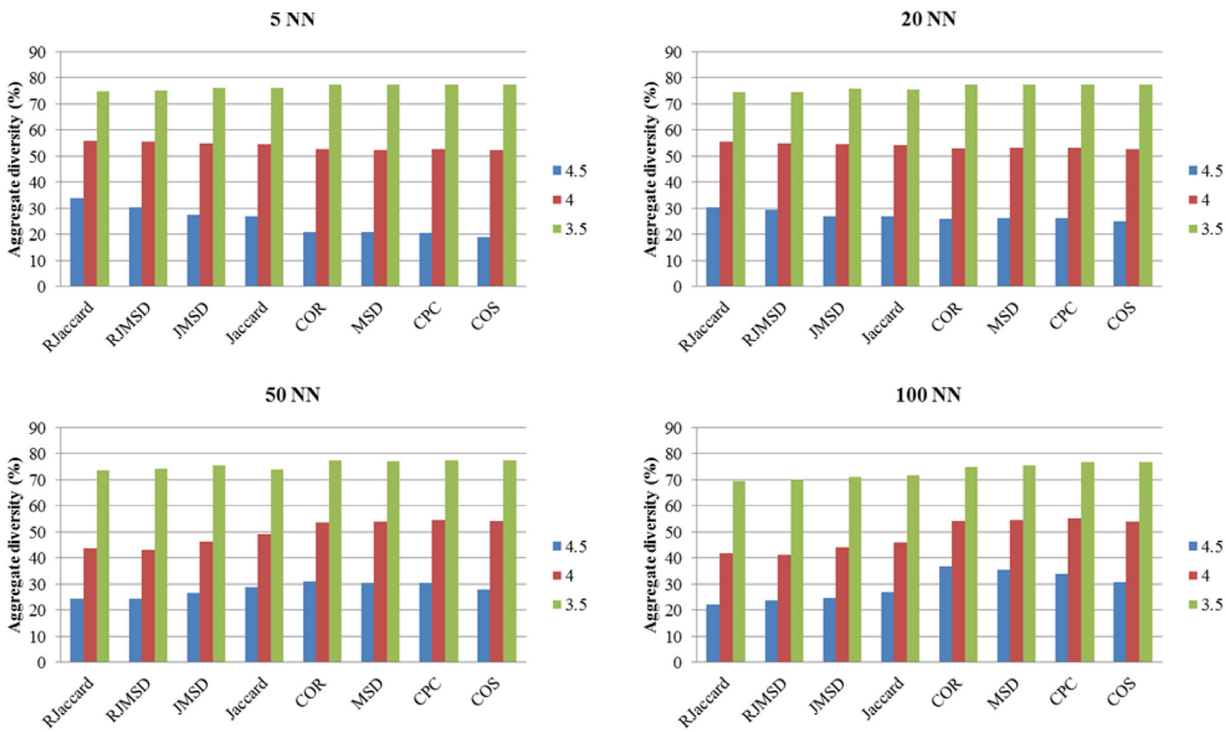


Fig. 4. Aggregate diversity of collaborative filtering using various traditional similarities.

## 6. Discussion and conclusion

### 6.1. Discussion

The goal of this research is to develop new simple but effective similarity models for improving the recommendation accuracy in lower computation time. To fulfill this objective, two new similarity metrics such as Relevant Jaccard Similarity and Relevant Jaccard Mean Square Distance have been developed in this study. Apart from developing two new simple but effective similarity methods, this research has a significant contribution to practice in the following ways. The proposed method identifies the relevant Jaccard similarity which helps to select appropriate nearest neighbors in the prediction model. Appropriate nearest neighbors benefit to recommender system to predict accurate ratings of unrated items. As a result, proposed RJaccard and RJMSD methods perform significantly well in a limited number of nearest neighbors. Furthermore, all rating vectors instead of only co-rated items are considered in this study for calculating similarity metrics. Therefore, proposed methods can also perform better in the cold start problem. Moreover, developed similarity metrics can be easily installed in existing collaborative filtering methods which are widely applied in online business areas of travel, e-retailing, online broadcasting, scientific articles, advertising, news, movie, and music etc.

### 6.2. Conclusion

In this study, two new similarity metrics such as Relevant Jaccard Similarity (RJaccard) and Relevant Jaccard Mean Square Distance (RJMSD) have been developed and tested for appropriate movie recommendations. Further, results of using RJaccard and RJMSD in recommender system model are compared with others published similarity measures in recommender system domain (i.e. Cosine Similarity (COS), Pearson's Correlation (COR), Constrained Pearson's Correlation (CPC), Jaccard similarity, Mean Square Distance (MSD), Jaccard Mean Square Distance (JMSD), Weighted Regularized Matrix Factorization (WRMF), and Factored Item Similarity Models (FISM) etc.). It is observed that RJaccard and RJMSD always outperform than other similarity approaches in terms of F1-Measure and aggregate diversity. It signifies that the profit of e-retailer and customer's satisfaction can be increased by applying proposed similarity metrics in the recommendation system. In another observation, it is noticed that RJaccard and RJMSD perform remarkably well while applying minimum nearest neighbors (i.e. 5). Hence, it indicates that proposed similarity methods can generate the best outcomes in lower computation time.

Outcomes of this study have been shown effective and noticeable even though has some limitations which can be extended in many directions. For instance, various malicious and non-malicious noises present in the rating data which distort the quality of recommendations. In the future, proposed similarity models can be incorporated with noise removal or noise correction methods [6] to formulate advance similarity approaches and solve noise rating related problems in the recommendation system. Moreover, proposed similarity models have not been associated with the contents of the user profile. Thus, in the future, hybrid collaborative filtering can be developed with the help of proposed similarity measures. Besides, this study of prediction for unrated items has been performed based on a $k$-nearest neighborhood recommendation system. In the future, other machine learning/extreme learning machine based algorithms can be applied to predict unrated items and various recommendations techniques can be compared to validate results. Furthermore, newly developed relevant Jaccard similarity model is performed only on a movie recommendation system. In the future, proposed similarity models can be applied in many application and research domain in replace of well-known Jaccard similarity approach.

## Acknowledgments

The work has been supported by the project of E-Business Center of Excellence and funded by Ministry of Human Resource and Development (MHRD), Government of India under the scheme of Center for Training and Research in Frontier Areas of Science and Technology (FAST), Grant No. F.No.5-5/2014-TS.VII.

## References

[1] G. Adomavicius, Y. Kwon, Improving aggregate recommendation diversity using ranking-based techniques, Knowl. Data Eng. IEEE Trans. 24 (2012) 896–911, doi:10.1109/TKDE.2011.15.
[2] G. Adomavicius, Y. Kwon, Optimization-based approaches for maximizing aggregate recommendation diversity, INFORMS, J. Comput. 26 (2014) 351–369, doi:10.1287/ijoc.2013.0570.
[3] H.J. Ahn, A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem, Inf. Sci. 178 (2008) 37–51, doi:10.1016/j.ins.2007.07.024.
[4] T.T. Allen, Z. Sui, N.L. Parker, Timely decision analysis enabled by efficient social media modeling, Decis. Anal. 14 (2017) 250–260, doi:10.1287/deca.2017.0360.
[5] S. Bag, M.K. Tiwari, F.T.S. Chan, Predicting the consumer's purchase intention of durable goods: an attribute-level analysis, J. Bus. Res. 94 (2019) 408–419, doi:10.1016/j.jbusres.2017.11.031.
[6] S. Bag, S. Kumar, A. Awasthi, M.K. Tiwari, A noise correction-based approach to support a recommender system in a highly sparse rating environment, D. Sup. Sys. (2019), doi:10.1016/j.dss.2019.01.001.
[7] J. Bobadilla, F. Serradilla, A. Hernando, Collaborative filtering adapted to recommender systems of e-learning, Knowledge-Based Syst. 22 (2009) 261–265, doi:10.1016/j.knosys.2009.01.008.
[8] K. Bok, J. Lim, H. Yang, J. Yoo, Social group recommendation based on dynamic profiles and collaborative filtering, Neurocomputing 209 (2016) 3–13, doi:10.1016/j.neucom.2015.11.120.
[9] J. Chen, S. Feng, J. Liu, Topic sense induction from social tags based on non-negative matrix factorization, Inf. Sci. 280 (2014) 16–25, doi:10.1016/j.ins.2014.04.048.

[10] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, K. Li, A disease diagnosis and treatment recommendation system based on big data mining and cloud computing, Inf. Sci. 435 (2018) 124–149, doi:10.1016/j.ins.2018.01.001.
[11] A. Dey, M. Jenamani, J.J. Thakkar, Senti-n-gram: an n-gram lexicon for sentiment analysis, Expert Syst. Appl. 103 (2018) 92–105, doi:10.1016/j.eswa.2018.03.004.
[12] GroupLens, MovieLens datasets, 2017. https://grouplens.org/datasets/movielens/.
[13] M.R. Hamedani, S.-W. Kim, JacSim: an accurate and efficient link-based similarity measure in graphs, Inf. Sci. 414 (2017) 203–224, doi:10.1016/j.ins.2017.06.005.
[14] Y. He, Z. He, Y. Deng, P. Zhou, IFPBM s and their application to multiple attribute group decision making, J. Oper. Res. Soc. 67 (2016) 127–147, doi:10.1057/jors.2015.66.
[15] J. Herlocker, J.A. Konstan, J. Riedl, An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms, Inf. Retr. 5 (2002) 287–310, doi:10.1023/A:1020443909834.
[16] Y. Hu, Y. Koren, C. Volinsky, Collaborative filtering for implicit feedback datasets, in: Data Mining, 2008, in: ICDM'08. Eighth IEEE Int. Conf., 2008, pp. 263–272, doi:10.1109/ICDM.2008.22.
[17] H. Hwangbo, Y.S. Kim, K.J. Cha, Recommendation system development for fashion retail e-commerce, Electron. Commer. Res. Appl. 28 (2018) 94–101, doi:10.1016/j.elerap.2018.01.012.
[18] S. Jiang, Q. Xueming, S. Jialie, F. Yun, M. Tao, Author topic model-based collaborative filtering for personalized POI recommendations, IEEE Trans. Multimedia 17 (6) (2015) 907–918, doi:10.1109/TMM.2015.2417506.
[19] S. Kabbur, X. Ning, G. Karypis, Fism: factored item similarity models for top-n recommender systems, in: Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2013, pp. 659–667, doi:10.1145/2487575.2487589.
[20] V.R. Kagita, A.K. Pujari, V. Padmanabhan, S.K. Sahu, V. Kumar, Conformal recommender system, Inf. Sci. 405 (2017) 157–174, doi:10.1016/j.ins.2017.04.005.
[21] J.B. Kim, P. Albuquerque, B.J. Bronnenberg, The probit choice model under sequential search with an application to online retailing, Manage. Sci. 63 (2016) 3911–3929, doi:10.1287/mnsc.2016.2545.
[22] F. Kong, J. Li, Z. Lv, Construction of intelligent traffic information recommendation system based on long short-term memory, J. Comput. Sci. 26 (2018) 78–86, doi:10.1016/j.jocs.2018.03.010.
[23] J. Lacasta, F.J. Lopez-Pellicer, B. Espejo-García, J. Nogueras-Iso, F.J. Zarazaga-Soria, Agricultural recommendation system for crop protection, Comput. Electron. Agric. 152 (2018) 82–89, doi:10.1016/j.compag.2018.06.049.
[24] X. Li, X. Cheng, S. Su, S. Li, J. Yang, A hybrid collaborative filtering model for social influence prediction in event-based social networks, Neurocomputing 230 (2017) 197–209, doi:10.1016/j.neucom.2016.12.024.
[25] G. Li, W. Ou, Pairwise probabilistic matrix factorization for implicit feedback collaborative filtering, Neurocomputing 204 (2016) 17–25, doi:10.1016/j.neucom.2015.08.129.
[26] C. Li, Z. Wang, S. Cao, L. He, WLRRS: A new recommendation system based on weighted linear regression models, Comput. Electr. Eng. 66 (2018) 40–47, doi:10.1016/j.compeleceng.2018.02.005.
[27] C. Lin, R. Xie, X. Guan, L. Li, T. Li, Personalized news recommendation via implicit social experts, Inf. Sci. 254 (2014) 1–18, doi:10.1016/j.ins.2013.08.034.
[28] G. Linden, B. Smith, J. York, Amazon.com recommendations: item-to-item collaborative filtering, IEEE Internet Comput. 7 (2003) 76–80, doi:10.1109/MIC.2003.1167344.
[29] H. Liu, Z. Hu, A. Mian, H. Tian, X. Zhu, A new user similarity model to improve the accuracy of collaborative filtering, Knowledge-Based Syst. 56 (2014) 156–166, doi:10.1016/j.knosys.2013.11.006.
[30] C.-L. Liu, X.-W. Wu, Fast recommendation on latent collaborative relations, Knowledge-Based Syst. 109 (2016) 25–34, doi:10.1016/j.knosys.2016.06.016.
[31] P. Melin, O. Castillo, J. Kacprzyk, in: Nature-Inspired Design of Hybrid Intelligent Systems, Studies in Computational Intelligence, Springer, 2017, p. 667, doi:10.1007/978-3-319-47054-2.
[32] F. Meng, Q. An, X. Chen, A consistency and consensus-based method to group decision making with interval linguistic preference relations, J. Oper. Res. Soc. 67 (11) (2016) 1419–1437, doi:10.1057/jors.2016.28.
[33] R. Ojha, A. Ghadge, M.K. Tiwari, U.S. Bititci, Bayesian network modelling for supply chain risk propagation, Int. J. Prod. Res. 56 (17) (2018) 5795–5819, doi:10.1080/00207543.2018.1467059.
[34] W. Pan, L. Chen, Group Bayesian personalized ranking with rich interactions for one-class collaborative filtering, Neurocomputing 207 (2016) 501–510, doi:10.1016/j.neucom.2016.05.019.
[35] T.K. Paradarami, N.D. Bastian, J.L. Wightman, A hybrid recommender system using artificial neural networks, Expert Syst. Appl. 83 (2017) 300–313, doi:10.1016/j.eswa.2017.04.046.
[36] Y. Park, J. Oh, H. Yu, RecTime: real-time recommender system for online broadcasting, Inf. Sci. 409 (2017) 1–16, doi:10.1016/j.ins.2017.04.038.
[37] B.K. Patra, R. Launonen, V. Ollikainen, S. Nandi, A new similarity measure using Bhattacharyya coefficient for collaborative filtering in sparse data, Knowledge-Based Syst. 82 (2015) 163–177, doi:10.1016/j.knosys.2015.03.001.
[38] N. Polatidis, C.K. Georgiadis, A dynamic multi-level collaborative filtering method for improved recommendations, Comput. Stand. Interfaces 51 (2017) 14–21, doi:10.1016/j.csi.2016.10.014.
[39] L. Ren, W. Wang, An SVM-based collaborative filtering approach for Top-N web services recommendation, Futur. Gener. Comput. Syst. 78 (2) (2018) 531–543, doi:10.1016/j.future.2017.07.027.
[40] M. Reusens, W. Lemahieu, B. Baesens, L. Sels, Evaluating recommendation and search in the labor market, Knowledge-Based Syst. 152 (2018) 62–69, doi:10.1016/j.knosys.2018.04.007.
[41] D. Sánchez, P. Melin, O. Castillo, A grey Wolf optimizer for modular granular neural networks for human recognition, Comput. Intell. Neurosci. (2017) (2017) 1–26, doi:10.1155/2017/4180510.
[42] B. Sarwar, G. Karypis, J. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: Proc. 10th Int. Conf. World Wide Web, 2001, pp. 285–295, doi:10.1145/371920.372071.
[43] A. Shahmohammadi, E. Khadangi, A. Bagheri, Presenting new collaborative link prediction methods for activity recommendation in Facebook, Neurocomputing 210 (2016) 217–226, doi:10.1016/j.neucom.2016.06.024.
[44] X. Shi, X. Luo, M. Shang, L. Gu, Long-term performance of collaborative filtering based recommenders in temporally evolving systems, Neurocomputing 267 (2017) 635–643, doi:10.1016/j.neucom.2017.06.026.
[45] S. Singh, S. Bag, M. Jenamani, Relative similarity based approach for improving aggregate recommendation diversity, in: In Proc. 12th Ieee Indicon Conf., 2015, doi:10.1109/INDICON.2015.7443856.
[46] X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. (2009) (2009) 1–19, doi:10.1155/2009/421425.
[47] T. Tran, D. Phung, S. Venkatesh, Collaborative filtering via sparse Markov random fields, Inf. Sci. 369 (2016) 221–237, doi:10.1016/j.ins.2016.06.027.
[48] Y. Wang, J. Deng, J. Gao, P. Zhang, A hybrid user similarity model for collaborative filtering, Inf. Sci. 418 (2017) 102–118, doi:10.1016/j.ins.2017.08.008.