Improved Collaborative Filtering Recommender System Based on Hybrid Similarity Measures

Mohamed Abdi
Department of Information Technology
Jomo Kenyatta University of Agriculture and Technology, Kenya
mhabdi@gmail.com

George Okeyo
Department of Information and Communications Technology, Carnegie Mellon University Africa, Rwanda
gokeyo@andrew.cmu.edu

Ronald Mwangi
Department of Information Technology
Jomo Kenyatta University of Agriculture and Technology, Kenya
waweru_mwangi@icsit.jkuat.ac.ke

Abstract: *Recommender Systems (RS) based on collaborative filtering has been successfully applied to provide relevant and personalized recommendations from an enormous amount of data in various domains. To achieve this, similarity measurements, such as the Pearson Correlation Coefficient (PCC), Cosine, and Jaccard, are used to compute the similarity between users or items based on correlations among user preferences from the user-item rating matrix. However, existing similarity metrics suffer from drawbacks emanating from data sparsity caused by insufficient number of transactions and feedback and scalability of the system's ability to handle increasing amounts of data efficiently. The objective of this study is to improve the recommendation quality and increase the prediction accuracy by addressing the problems of similarity computation in collaborative filtering. This paper presents a hybrid similarity measure that combines Adjusted Triangle similarity, User Rating Preference behavior, and the Jaccard (ATURPJ) coefficient. The proposed hybrid similarity measures were evaluated on four widely used and publicly available datasets, MovieLens, FilmTrust, and CiaoDVD, using the predictive accuracy metrics of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) and recommendation quality of Precision, Recall, and F-measure. The experimental results show that the proposed hybrid similarity measure significantly outperforms existing approaches with MAE of 0.547 and RMSE of 0.735 compared to the baseline of 0.707 and 0.903 respectively on ML-100k dataset. Overall, this approach has the potential to improve the quality of recommendation and accuracy of the prediction.*

Keywords: *Recommender systems, collaborative filtering, similarity measure, Adjusted Triangle similarity, Jaccard similarity, user rating preference behavior.*

Received May 1, 2024; accepted October 14, 2024
https://doi.org/10.34028/iajit/22/1/8

1. Introduction

The advent of new technological innovations has accelerated the rate of production, sources and dissemination of information. We are continuously bombarded with information on daily basis from different and varied channels. It has become increasing difficult to find the required information quickly and conveniently from various print, electronic and online sources. Information overload, characterized by an over-abundance of information is a major cause of concern for general information users, researchers and business managers. The social media, in particular, with user generated content across the platforms has also contributed to the problem of information overload [31]. In fact, Roetzel [42] opines that information overload can be seen as a virus, spreading through (social) media and news networks. Since its inception in 2005, YouTube has continued to offer abundance of choice; 300 hours of video are uploaded every minute, gets over 30 million visitors per day and financially, it's a $15 billion-a-year business [15]. Timely, accurate and relevant information is crucial to individuals, businesses and governments and all sectors of the economy for decision making. Personalization has become one of the key features of online content.

Recommender Systems (RS) are crucial tools to overcome the information overload brought about by the advent of the internet, the ongoing digitalization of the world of work and the growing use of information and communication technologies [6]. RS are a class of web applications that assist users to tame the problem of information overload by providing personalized recommendations on various types of products and services. An example is a RS offering news articles to online newspaper readers, based on a prediction of reader interests or that of an online retailer giving suggestion about what they might like to buy, based on their past history of purchases and/or product searches. RS recommendation approaches are generally classified [51] into content-based, collaborative filtering, knowledge-based and hybrid filtering. Content-based approaches attempt to build a user profile to predict ratings on unseen items while knowledge-based uses explicit knowledge about products and users to create a knowledge-based criterion to generate recommendations. Hybrid systems combine two or more techniques to obtain better performance [12]. There are several properties of recommender system that can be evaluated, for example accuracy, novelty, diversity and serendipity [29]. Recommendation accuracy and quality are quantified using error metrics like Mean Absolute Error (MAE), Root Mean Square Error (RMSE), or quality metrics like Precision, Recall and F-measure through offline or online studies of algorithm's performance [52]. The deployment of RS has alleviated the problem of information overload by helping users find relevant products and services based on user preferences and constraints. Over the years, the techniques and applications of RSs have evolved in both research and industry because of the exponential growth in the number of online users and content.

1.1. Collaborative Filtering

Collaborative Filtering (CF) is one of the most popular and widely used successful methods for recommendation systems because of its simplicity and efficiency [49]. The CF method recommends items that are similar to items previously preferred by a specific user. The basic principles of Collaborative Filtering-Recommender Systems (CF-RS) are to:

a) Analyze the description of the items preferred by a particular user to determine the principal common attributes that can be used to distinguish these items and store them in a user profile.
b) Compare each item's attributes with the user profile so that only items that have a high degree of similarity with the user profile will be recommended [16].

The CF approach is further categorized into memory-based and model-based approaches [14] Memory-based CF algorithms utilize the entire user-item database to generate a prediction. These systems employ statistical techniques to find a set of users, known as neighbors, that have a history of agreeing with the target user. As described in Breese *et al.* [11], the memory-based method first calculates the similarities among users and then selects the most similar users as neighbors of the active user and predict recommendations according to nearest neighbors. The memory-based approach can provide considerable recommended accuracy compared to the model-based approaches [7]. The model-based method first constructs a model to describe the behavior of users, and therefore, to predict the ratings of items. It provides item recommendations by first developing model of user ratings. Algorithms in this category take a probabilistic approach and envision the collaborative filtering process as computing the expected value of a user prediction, given the user ratings on other items [43].

The computational efficiency of both approaches depends on the ratio between the number of users and items. The memory-based method is adaptive to data changes, but requires large computational time according to the data size. This means as the number of users and items increases, the computational time will escalate significantly. As for model-based method, it has a constant computing time regardless the size of the data but not adaptive to data changes. The memory-based CF is commonly leveraged by online companies because it is efficient and easier to be implemented than the model-based CF [5].

To create a working recommender system, a substantial amount of data must be collected in the form of user ratings for items in a user-item matrix. Collaborative filtering algorithms enables users to rate a collection of elements (such as videos, songs, or movies, in a CF based website) so that when sufficient data are stored on the system, recommendations can be made to each user based on the information provided by the users we consider to share the most in common with them [10].

The CF-based approach often suffers from several shortcomings [13], such as data sparsity, cold start, and scalability issues, which seriously affect the recommended quality of RS. The sparsity problem occurs since not all users rate all items, which reduces the amount of data available to the CF. If we consider every user as a vector of ratings on items, and we put them in a matrix, the resulting user-item matrix, the traditional input of CF techniques, is very sparse. Sparsity refers to irregular, insufficient, or highly variable user rating. The cold start problem in memory-based CF arises when there is no rating available when a new user or item enters the system. As the number of users and items continues to grow, the user-item matrix becomes increasingly large, [45] and the CF encounters scalability limitations, leading to performance degradation, inaccurate predictions, and below standard recommendations.

1.2. Similarity Metrics

Similarity computations play an important role in the recommendation process. Historically, a number of statistical measures, including Pearson Correlation Coefficient (PCC), Jaccard Index (JI), Cosine Correlation Coefficient (COS), and Mean Squared Differences (MSD), have been applied [36]. The existing similarity measures primarily depend on the common items rated by users. Pearson [40] the similarity is computed by considering the co-rated items and if the co-rated items are few, then the performance of PCC is impaired, leading to inaccurate predication. JI [24] produces result as absolute values either 0.5 or 1.0. It is therefore difficult to differentiate users when the output return is only absolute values. The cosine between two vectors representing two users or items indicates the similarity value between each other. A value close to 1 indicates the existence of a strong correlation between the two variables while a value close to 0 indicates that there is no correlation. The Cosine similarity measures does not consider the users rating preference. Absolute ratings are taken into account via MSD rather than the total amount of standard ratings. Similarity is determined by averaging these squared differences; the smaller the MSD, the more similar two users or items are [18]. These limitations are inherent in all the existing similarity metrics and cannot be used in sparse rating environment as the similarity computation leads to inaccurate recommendation to the users.

1.3. Hybrid Similarity Measures

Metrics have recently been developed to overcome the limitations of CF to provide accurate predictions and quality recommendations [9] by attempting to address the sparsity, cold start and scalability challenges in CF recommender system. The primary focus is to tackle the insufficient ratings in the user-item rating matrix that presents challenges to effectively identify similar users or items to model the user preferences. This can be achieved by formulating new similarity measure that can sufficiently utilize the limited available rating information. To address these problems, a new hybrid similarity measure, ATURPJ, which integrates Adjusted Triangle, User Rating Preferences behavior and Jaccard similarity (ATURPJ) is proposed in this paper.

The main goal of this study is to devise a similarity computation method that works for collaborative filtering regardless of the sparsity of the datasets. This is anticipated to increase the prediction accuracy and improve the quality of recommendations by addressing the challenges of data sparsity and scalability in collaborative filtering.

This paper proposes to improved collaborative filtering recommender system based on a hybrid similarity measure integrating ATURPJ behavior and the Jaccard similarity measures. The main contributions of this study are summarized as follows:

a) We formulated an Adjusted Triangle similarity measure that considers both the co-rated items and the ratings of items rated by either user. This implies that the model utilizes all available rating data to compute the similarity measure, thus alleviating the data sparsity problem.
b) To reflect the variation in the rating preference of users, we introduced the mean and standard variance of user ratings to complement the Adjusted Triangle similarity measure.
c) The proposed improved hybrid similarity measure of ATURPJ is based on the Adjusted Triangle similarity, user rating preferences behavior, and Jaccard similarity.
d) We conducted comprehensive experiments to evaluate the performance of our approach and compared it with the baseline similarity measure based on publicly available datasets.

The remainder of this paper is organized as follows. In section 2, related work on the similarity measure and its drawbacks are presented and discussed. Section 3 outlines the details of the proposed hybrid similarity measure and similarity measurement process. Section 4 presents the results, analysis, and comparison with baseline methods. Finally, section 5 presents the conclusions and future work.

2. Related Work

To effectively overcome the challenges posed by the sparseness of data and the drawbacks of traditional similarity measures in memory-based and model-based collaborative filtering, researchers have proposed a variety of techniques and hybrid similarity metrics to improve the performance of the recommender system, accuracy of the prediction and the quality of recommendation. Numerous studies that propose new, hybrid, or modified similarity measurements have been published in the literature to enhance CF performance. The recent studies focused on hybrid similarity measures, in recognition of the fact that the conventional measures have shortcomings that degrade its performance in sparse data environments or where the cold start problem arises.

Liu *et al*. [34] proposed a similarity measure known as the New Heuristic Similarity Method (NHSM). It computes three parameters for each co-rated item: Proximity, Significance, and Singularity (PSS). Subsequently, each computed parameter is multiplied by the modified Jaccard similarity. The obtained similarity is again multiplied with URP function to obtain the resultant NHSM similarity. Although an improved similarity metric is reported, the major drawback of this approach is that the computation of NHSM similarity is complex and lengthy, which makes it difficult for the recommender system to produce results in real time. All factors in NHSM are multiplied repeatedly, which ultimately weakens the performance, and combining these results with other similarity measures becomes complex.

The study by Suryakant and Mahara [48] combined Cosine, Jaccard and Mean Measure of Divergence (CjacMD) to compute the overall similarity for evaluating sparse datasets. The Mean Measure of Divergence (MMD) was proposed to take into account the rating habits of a user. The CjacMD is disadvantaged by the shortcomings of Jaccard and the cosine similarity.

By integrating Triangle and Jaccard similarities (TMJ) for recommendation [47], the accuracy of prediction was improved by complementing Triangle and Jaccard similarities. Triangle and Jaccard hybridizations are designed to provide information on co-rated as well as non-co-rated users. However, the study limitation emanates from the Jaccard and Triangle similarity measure, which do not consider the absolute value ratings, ignore the co-rated items and the lack of incorporation of user rating preference respectively.

The factor-based approach [50] is another attempt to enhance the accuracy of CF by incorporating the user rating average, user rating variance, and number of overlapping ratings into the measurement of user similarity. The aim is to strengthen the role of individual users who are genuinely similar to the active user on item preferences, while reducing the weight of users who are different from the active user in terms of rating patterns. The study investigated using the PCC only and therefore the performance of the factor approach with other similarity measures cannot be inferred from the study.

In a similar study, Alshammari *et al*. [4] proposed a Triangle multi-level item-based collaborative filtering approach with a Triangle similarity measure that considers the length and angle of rating vectors between users and allows positive and negative adjustments using a multi-level recommendation technique. While the study reported improved accuracy PCC and multi-level CF, there is need to conduct a comparative study with hybrid similarity metrics and using different datasets with different sparsity levels. A comparative study of seven commonly used similarity measures showed that the weighted PCC had the best prediction accuracy when the dataset was sparse [2]. The study used only predictive accuracy metrics and on a very small dataset. The study by Feng *et al*. [17], propose similarity model composed of three impact factors of S1, S2, and S3. S1 expresses the similarity between users, S2 calculates the number of co-rated items to be less than the specified threshold, and S3 explains the weight of each user rating value. The time complexity of the item similarity computation for the similarity impact factors is unknown, and therefore its suitability in real time. Bag *et al*. [8] proposed two similarity measurement models, Relevant Jaccard (RJaccard) and Relevant Jaccard Mean Square Distance (RJMSD), both generated by multiplying two similarity metrics and exploring the use of all scores and constructing a similarity computation model to overcome the challenge of using only commonly rated items or users in Jaccard similarity computation. In this study, the relevant Jaccard similarity was to select appropriate nearest neighbors in the prediction model and the RJMSD is used in prediction of unrated items of each user. Amer *et al*. [5] designed three similarity metrics of Difference-based Similarity Measure (SMD), Hybrid Difference-Based Similarity Measure (HSMD), and Triangle-based cosine measure (TA) to tackle the data sparsity problem. The SMD and TA were proven to be superior comparative analysis. The reported results for TA, MSD, HSMD and HSMDJ are very marginal.

A Modified Proximity-Impact-Popularity (MPIP) version of Proximity-Impact-Popularity (PIP) was proposed by Manochandar and Punniyamoorthy [35], to address the high range of values for each component in PIP by fixing the range from 0 to 1. A new similarity method Efficient Gowers-Jaccard-Sigmoid Measure (EGJSM) based on the Efficient Gower's (EG), Jaccard, and efficient PSS methods is proposed by Jain *et al*. [26]. It is observed that the performance of MPIP measure deteriorates when the dataset contains cold-start users. To deal with this issue, Jain *et al*. [26] proposed a new coefficient named efficient PSS by combining efficient PSS factors. In efficient PSS, the efficient proximity coefficient is developed based on the proximity factor of PIP [35]. The study by Abdalla *et al*. [1] focused on five new item-based similarity measures to address the cold-start and data sparsity problems based on the ideologies of the Term Frequency (TF) and Inverse Document Frequency (IDF), Question and Test Interoperability (QTI) and Question and Test Interoperability Jaccard (QTIJ), combination of COSinE and Jaccard (COSEJ) and Numerical Proximity Similarity Measure (NPSM). The study did not elucidate on how the limitations of Jaccard are handled in the proposed hybrid similarity measures.

While all the studies reported enhanced accuracy, none of the proposed approaches addressed the inherent limitations of the Cosine, Jaccard, Pearson and Triangle similarity measures. The Jaccard similarity computation process depends only on the proportion of common ratings, and does not consider the absolute value of the rating. Triangle similarity computation considers both the length and angle between two vectors and co-rated items but ignores items rated by either user [3]. Similarly, the Jaccard coefficient considers co-rated items, but ignores the absolute ratings and items rated by either user. In this case, both Triangle and Jaccard ignore the ratings of items by either user. The MSD considers the absolute rating differences among users to compute similarity. It fails to consider the proportion of common ratings and works only on co-rated items. Hence, the accuracy of the similarity value estimated by MSD depends on the number of co-rated items. The main limitation of cosine similarity is that its output has high/low similarity, despite significant similarities or differences in ratings [25]. Meissa *et al*. [38] proposed in their work an improved recommendation model which enables to discover personalized web Application Programming Interface (API) to provide personalized suggestions for users without sacrificing the recommendation accuracy.

3. Proposed Hybrid Similarity Metrics

3.1. Motivation

The CF suggests items to the user, based on the similarity between the user or the items. In CF algorithms, similarity metrics are the core components, their performances directly influence the accuracy of the prediction and the quality of the recommendations generated. The CF approach encounters data sparsity and scalability limitations as the volume of accessible information and the active users continues to grow leading to performance degradation, erroneous predictions and recommendations. To alleviate this problem, researchers have proposed many similarity measures for CF but they all suffer from drawbacks that lead to low or high similarity measure. The CF algorithm locates a user's neighbors defined as the users with rating history similar to the present user and then uses the ratings of the neighbors to produce recommendations. In most cases, similarity is calculated using a user-item rating matrix, where each row represents a rating vector evaluated by a related user, and each column represents the user ratings of a certain item. When deciding on the optimal algorithm to calculate the similarity, several restrictions must be considered. When algorithms are employed individually, they exhibit several shortcomings. For instance, while the Jaccard coefficient considers both the common items and the items that are present in either entity, cosine similarity algorithms only consider the common items that have been rated for measuring similarity.

Table 1 presents a hypothetical user-item rating matrix. We assume that there are five users and five items in the system. The symbol (-) represents the missing ratings in the user-item rating matrix.

Table 1. Example of a user-item rating matrix.

	Item1	Item2	Item3	Item4	Item5
User1	5	4	5	3	2
User2	4	2	-	-	4
User3	2	4	2	2	3
User4	4	3	-	1	2
User5	2	1	3	-	1

Table 2. Item-based similarity measure for different similarity metrics.

		i_1	i_2	i_3	i_4	i_5
Item-based similarity measure for Jaccard	i_1	1	1.000	0.600	0.600	1.000
	i_2	-	1	0.600	0.600	1.000
	i_3	-	-	1	0.500	0.600
	i_4	-	-	-	1	0.600
	i_5	-	-	-	-	1
		i_1	i_2	i_3	i_4	i_5
Item-based similarity measure for Triangle	i_1	1	0.777	0.916	0.655	0.721
	i_2	-	1	0.748	0.704	0.749
	i_3	-	-	1	0.776	0.622
	i_4	-	-	-	1	0.780
	i_5	-	-	-	-	1
		i_1	i_2	i_3	i_4	i_5
Item-based similarity measure for Adjusted Triangle	i_1	1	0.777	0.596	0.513	0.721
	i_2	-	1	0.638	0.645	0.749
	i_3	-	-	1	0.622	0.514
	i_4	-	-	-	1	0.533
	i_5	-	-	-	-	1
		i_1	i_2	i_3	i_4	i_5
Item-based similarity measure for TMJ	i_1	1	0.777	0.596	0.513	0.721
	i_2	-	1	0.638	0.645	0.749
	i_3	-	-	1	0.622	0.514
	i_4	-	-	-	1	0.533
	i_5	-	-	-	-	1
		i_1	i_2	i_3	i_4	i_5
Item-based similarity measure for ATURP	i_1	1	0.777	0.358	0.308	0.721
	i_2	-	1	0.383	0.387	0.749
	i_3	-	-	1	0.311	0.308
	i_4	-	-	-	1	0.320
	i_5	-	-	-	-	1
		i_1	i_2	i_3	i_4	i_5
Item-based similarity measure for ATURPJ	i_1	1	0.384	0.179	0.114	0.328
	i_2	-	1	0.187	0.167	0.364
	i_3	-	-	1	0.112	0.138
	i_4	-	-	-	1	0.153
	i_5	-	-	-	-	1

Triangle similarity considers the length and angle of the rating vectors between them, as well as the common ratings of users. However, Triangle similarity:

a) Ignores items that are rated by either user.
b) Does not consider user rating preference.

The Jaccard similarity measure is given by the ratio of the intersection to the union between items, as given in Equation (11). It considers items that are commonly rated by users, but it does not include the absolute value of the user's ratings. We address these shortcomings by applying similarity-different computations to those listed in Table 1. The results of the computations for the item-based similarity measure using different similarity metrics are shown in Table 2.

The following are some of the deficiencies of the Triangle similarity measure that we seek to address in this study:

a) The Triangle similarity between item1 and item2 is computed as 0.777 while between Item1 and item3 is 0.916, which is higher. This is misleading since item1 and item2 have very similar ratings. Furthermore, the total rated values of item2 were higher than item3. This is because Triangle similarity does not consider ratings that are not rated by common users. The Adjusted Triangle similarity between item1 and item3 is 0.596, which is lower. The Adjusted Triangle similarity method considers the ratings of items not rated by common users. The proposed similarity measures of ATURP and ATURPJ were 0.298 and 0.179, respectively. The ATURPJ provides the most reliable and accurate similarity measure. The TMJ suffers from inherent.
b) The TMJ similarity measure, which is a product of a Triangle and Jaccard, suffers from the same problem of not considering the ratings of items not rated by common users. To illustrate this further, the TMJ similarity between item2 and item3 was computed as 0.383, whereas ATURP and ATURPJ, which used the Adjusted Triangle similarity, were 0.312 and 0.187, respectively. Table 3 provides a summary of these comparisons.

Table 3. Summary of similarity computations with different similarity measure between items.

Similarity between	Triangle	Adjusted Triangle	TMJ	ATURP	ATURPJ
Item1 and item 3	0.916	0.596	0.358	0.298	0.179
Item1 and item 4	0.655	0.513	0.308	0.189	0.114
Item2 and item 3	0.748	0.638	0.383	0.312	0.187
Item2 and item 4	0.704	0.645	0.387	0.276	0.167
Item3 and item 5	0.622	0.514	0.308	0.153	0.138
Item4 and item 5	0.780	0.533	0.320	0.256	0.153

The similarity measure model employed for CF-RS is the hybrid integration of ATURPJ. This paper presented ATURPJ, a hybrid similarity measure that integrates Triangles, User Rating Preference behavior, and Jaccard similarity metrics. The proposed method was experimentally evaluated using four widely used datasets. These hybrid methods are used to enhance the effectiveness of RS by considering CF issues. We were motivated to address the problems of Triangle similarity and complement it with URP and Jaccard similarity to obtain more accurate hybrid similarity measures. Empirical evidence shows that information overload is positively related to strain, burnout and various health related complications [37]. In a healthcare setting, it has been associated with higher error rates which negatively impact patient safety [39]. The proposed hybrid similarity metric is expected to reduce the burden of information overload by providing accurate, reliable and relevant recommendations.

3.2. Formulation of the Hybrid Similarity Metric

3.2.1. Symbols and Notations

The basic symbols and notations to describe different concepts and user-item lists with similar interests and preferences are presented in Table 4.

Table 4. Table of basic notations.

Symbol	Descriptions/Notations
U	Set of users
I	Set of items to be rated
m	Number of users, $\|U\|$
n	Number of items, $\|I\|$
u	User $u \in U \wedge u \leq m$
v	User $v \in U \wedge v \leq m$
i	Item $i \in I \wedge i \leq n$
k	Size of user neighbourhood
ϵ	Similarity threshold for inclusion in user neighbourhood
N	Number of highest rated items to recommend
I_u	Set of items rated by user $u \in U$
I_v	Set of items rated by user $v \in U$
$I_{u,v}$	Set of items rated by at least one of the two users $u \in U$ or $v \in U$
$R_{m \times n}$	$m \times n$ user-item rating matrix
$S_{m \times m}$	$m \times m$ user-user similarity matrix
$P_{m \times n}$	$m \times n$ user-item predicted rating matrix
$R_{u,i}$	User u rating for item i
$R_{v,i}$	User v rating for item i
$S_{u,v}$	Similarity score between user u and v
$P_{u,i}$	Predicted rating for item $i \in I$ by user $u \in U$
$\bar{R}_u$	Mean rating of user $u \in U$
R_u	Vector of user ratings $(R_{u,i_1}, R_{u,i_2}, .. R_{u,i_n}) \in R_{m \times n}$, by user u, for items $i_1, i_2, ...i_n \in I$
R_v	Vector of user ratings $(R_{v,i_1}, R_{v,i_2}, .. R_{v,i_n}) \in R_{m \times n}$, by user v, for items $i_1, i_2, ...i_n \in I$
V_k	List of k nearest neighbours, $V_1, V_2, \ldots, V_k \in U$, of user u, such that, $S_{u,V1} > S_{u,V2} > \cdots S_{u,Vk} > \epsilon$
sim(u, v)	Choice of similarity measure between user u and v: $sim^{TRIANGLE}(u,v)$, $sim^{ATRIANGLE}(u,v)$, $sim^{Jaccard}(u,v)$, $sim^{ATURPJ}(u,v)$

3.3. Similarity Metrics

3.3.1. Triangle and Adjusted Triangle Similarity

The Triangle similarity measure suffers from two major limitations:

1. It works only on co-rated items and ignores items rated by either user.
2. It does not consider URP. We have addressed these two limitations as follows

- **Definition 1: Triangle Similarity**

The Triangle similarity (T) between the rating vectors vector R_u and vector R_v, of users u and v is defined as follows:

$$Sim(u,v)^T = 1 - \frac{\sqrt{\sum_{i \in I_{u,v}} (R_{u,i} - R_{v,i})^2}}{\sqrt{\sum_{i \in I_{u,v}} R_{u,i}{}^2} + \sqrt{\sum_{i \in I_{u,v}} R_{v,i}{}^2}} \quad (1)$$

Given the set $I_{u,v}$ of items, rated by both users, that is, for any two users u and v,

If $I_u=\{i|R_{u,i}>0\}$ and $I_v=\{i|R_{v,i}>0\}$ are the set of items rated by user u and v respectively, then, $I_{u,v}=I_u \cap I_v$.

From geometry, Equation (1) can be rewritten as follows:

$$T(\overrightarrow{0A}, \overrightarrow{0B}) = 1 - \frac{|AB|}{|0A| + |0B|} = 1 - \frac{|R_u - R_v|}{|R_u| + |R_v|} \quad (2)$$

where $\overrightarrow{0A}$ is the rating vector R_u and $\overrightarrow{0B}$ is the rating vector of R_v, for user u and v, respectively, for items in $I_{u,v}$. The value of the triangular similarity range was [0,1] where 0 indicates no similarity, and the larger the value, the higher the similarity. The vectors $\overrightarrow{0A}$ and $\overrightarrow{0B}$ are taken as the sides of a Triangle, $\Delta\overrightarrow{0A}$, $\overrightarrow{0B}$, $\overrightarrow{AB}$ as illustrated in Figure 1. The rating vectors of two items form a Triangle in the space.

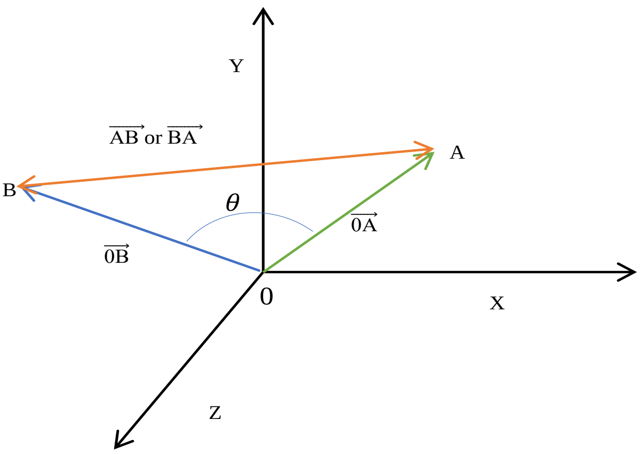

Figure 1. Geometric illustration of Triangle similarity measure.

- **Definition 2: Adjusted Triangle Similarity**

Let $I_{u,v}$ be the subset of item i, either rated by user u or v, that is, $I_{u,v}=\{i \in I_u \cup I_v\}$. The Adjusted Triangle similarity (AT) is given by:

$$Sim(u,v)^{AT} = 1 - \frac{\sqrt{\sum_{i \in I_{u,v}} (R_{u,i} - R_{v,i})^2}}{\sqrt{\sum_{i \in I_{u,v}} R_{u,i}{}^2} + \sqrt{\sum_{i \in I_{u,v}} R_{v,i}{}^2}} \quad (3)$$

The Adjusted Triangle similarity measure considers the co-rated items as well as the ratings of items rated by either user. The Triangle similarity computation considers both the length of the vectors and the angle between them, unlike cosine similarity, which depends only on the angle between the vectors. Geometrically, Triangle similarity and Adjusted Triangle similarity are equal, the difference between the two is that, in Adjusted Triangle similarity, $I_{u,v}$ is the set of items rated by the two users *u* and *v*. The Adjusted Triangle similarity measure considers the ratings of co-rated items as well as the ratings of items rated by either user.

3.3.2. Integrating Adjusted Triangle and URP

The user rating preference considers the rating preference of each user because different users have different rating behaviors. This is because some users generally prefer to give high ratings, whereas others prefer low ratings.

- **Definition 3: User Rating Preference (URP)**

To reflect this URP behavior, Liu *et al*. [34] proposed adopting the mean and deviation of the rating to model user preference. The URP similarity measure is defined as follows:

$$Sim(u,v)^{URP} = 1 - \frac{1}{1 + exp\left(-|\mu_u - \mu_v| \cdot |\sigma_u - \sigma_v|\right)} \quad (4)$$

where μ_u and μ_v denote the mean ratings of users *u* and *v* on item $i \in I_{u,v}$ respectively, whereas σ_u and σ_v represent the standard deviations of the users *u* and *v*, respectively. The mean and standard deviation were defined as follows:

$$\mu_u = \sum_{i \in I_u} \frac{R_{u,i}}{|I_u|} \quad (5)$$

$$\sigma_u = \sqrt{\sum_{i \in I_u} \frac{\left(R_{u,i} - \bar{R}_u\right)^2}{|I_u|}} \quad (6)$$

In providing preferences for items, each user has unique personal behavior that influences how rating preferences are expressed. Some users tend to generally give high ratings, whereas others give low ratings, irrespective of the prevailing circumstances. This introduces bias in the rating matrix, which influences the similarity between users and items. Traditional similarity does not consider this type of user behavior. It is desirable to capture different URP and integrate them into the similarity computation. Integrating the AT and URP behavior similarity measures to consider individual URP behavior. In Equation (7), (8), and (9), we combine AT, URP, and Jaccard to take advantage of both the existence and quantity available of rating values. The combination is mutual and not resonant; hence, we used a multiplicative combination of AT, URP, and Jaccard. In Equation (7) and (8), we combine *AT* and *URP*.

$$Sim(u,v)^{ATURP} = Sim_{(u,v)}^{AT} . Sim_{(u,v)}^{URP} \quad (7)$$

$$Sim(u,v)^{ATURP} = \left(1 - \frac{\sqrt{\sum_{i \in I_{u,v}} \left(r_{u,i} - r_{v,i}\right)^2}}{\sqrt{\sum_{i \in I_{u,v}} {r_{u,i}}^2} + \sqrt{\sum_{i \in I_{u,v}} {r_{v,i}}^2}}\right) . \left(1 - \frac{1}{1 + exp(-|\mu_u - \mu_v| . |\sigma_u - \sigma_v|)}\right) \quad (8)$$

3.3.3. Integrating ATURP with Jaccard

The Jaccard coefficient [33] mainly focuses on global ratings. It evaluates similarity as the ratio of the proportion of the cardinality of co-rated items to that of all items rated by users *u* and *v*, as defined in Equation (9).

$$Sim(u,v)^{Jaccard} = \frac{|I_u \cap I_v|}{|I_u \cup I_v|} \quad (9)$$

Where I_u and I_v are the sets of items rated by users *u* and *v*, respectively.

The Jaccard similarity index largely focuses on global ratings. This is the ratio of the proportion of the cardinality of co-rated items to the cardinality of all items rated by both users. Because I_u and I_v are not mutually exclusive, the addition rule theorem is applied to Equation (9) to obtain:

$$|I_u \cup I_v| = |I_u| + |I_v| - |I_u \cap I_v| \quad (10)$$

Where $|I_u|$+$|I_v|$ is the cardinality of sets I_u and I_v respectively. Therefore, similarity is computed as follows:

$$Sim(u,v)^{Jaccard} = \frac{|I_u \cap I_v|}{|I_u| + |I_v| - |I_u \cap I_v|} \quad (11)$$

The Jaccard similarity measure considers non-co-rating users in similarity computation but performs well in the similarity calculations of no common rating of users or items. We further combined the Jaccard similarity measure with *ATURPJ* to provide more information about non-co-rating users, thereby obtaining a new hybrid measure as follows:

$$\begin{aligned} Sim(u,v)^{ATURPJ} &= Sim_{(u,v)}^{ATURP} . Sim(u,v)^{Jaccard} \\ &= Sim_{(u,v)}^{ATURP} . \frac{|I_u \cap I_v|}{|I_u| + |I_v| - |I_u \cap I_v|} \end{aligned} \quad (12)$$

In this manner, we addressed the shortcomings of Triangle similarity by complementing it with URP and the Jaccard similarity measure.

3.4. Similarity Measurement Process

The complete workflow of our algorithm ATURPJ is presented in Algorithm (1). The following is a summary of the major steps:

- *Step* 1. Input.

In the input step, the users' profiles, items' profiles, and rating information are collected in the form of user, item ratings matrix. Five-fold cross validation on ratings dataset to obtain 20% test and 80% training set. Remove null data points; items with no rating and duplicate ratings.

- *Step* 2. ATURPJ hybrid similarity measure computation.

Compute the similarity between the user rating vectors and store the result in a similarity matrix.

- *Step* 3. Neighborhood generation.

Select the nearest neighbors of the target user using the k nearest neighbor based on their similarity.

- *Step* 4. Prediction and top *N* recommendations.

Compute the predicted rating based on Resnick's prediction formula by using Equation (13) and finally recommend top-N list of items.

- *Step* 5. Output.

Evaluation results using the accuracy and quality metrics.

A memory-based CF approach was used to predict the ratings. All similarity measures were tested using the item-based kNN algorithm on the given rating matrix. In this case, the kNN algorithm implies the user-based kNN and the rating matrix implies the item-based rating matrix. The predicted rating $P_{u,i}$ was computed using the Resnick's rating prediction formula [41] as follows:

$$P_{u,i} = \bar{R}_u + \frac{\sum_{v=1}^{k}(R_{v,i} - \bar{R}_v) \times sim(u,v)}{\sum_{v=1}^{k} sim(u,v)} \quad (13)$$

where k is the size of the neighborhood and $sim(u,v)$ computes the similarity between users u and v.

Algorithm (1) describes the pseudocode for the proposed hybrid similarity measurement in the recommendation process.

Algorithm 1: Procedure for CF top N Recommendation Using a Selected sim(u,v), Similarity Measure.

Input:
Set of users: U
Set of items: I
User-item rating matrix: $R_{m\times n}$
Size of neighborhood: k
Neighbor similarity threshold: ϵ
Number of top items: N
Output:
RMSE
MAE
PRECISION
RECALL
FMEASURE
Begin Procedure:
1. Initialize the similarity matrix $S_{m\times m}$
2. Initialize user-item predicted rating matrix $P_{m\times n}$
3. for each user u=0 to u=m do:
4. for each user v=0 to v=m do:
5. Using sim(u,v) compute the similarity between user rating vectors R_u *and* R_v *of the user-item rating matrix* $R_{m\times n}$*, of user u and v respectively.*
6. Store the result at $S_{u,v}$ *in similarity matrix* $S_{m\times m}$
7. end loop
8. for each item i=0 to i=n do:
9. Collect the list v_k *of k nearest neighbors for user u, whose similarity to user u is greater that* ϵ.
11. Predict the rating for item i by user u, using the Resnick CF recommender prediction formula.
12. Store the result at $P_{u,i}$ *in user-item predicted rating matrix* $P_{m\times n}$
13. end loop
14. end loop
15. Compute RMSE between the user-item rating matrix $R_{m\times n}$ *and the predicted rating matrix* $R_{m\times n}$
16. Compute MAE between the user-item rating matrix $R_{m\times n}$ *and the predicted rating matrix* $R_{m\times n}$
17. Compute PRECISION using the user-item rating matrix $R_{m\times n}$ *and the predicted rating matrix* $R_{m\times n}$
18. Compute RECALL using the user-item rating matrix $R_{m\times n}$ *and the predicted rating matrix* $P_{m\times n}$
19. Compute FMEASURE using PRECISION and RECALL
End Procedure

From the definition of our proposed model, the time complexity of user similarity computation, assuming an m (users)×n (items) matrix, is $O(n)$. In this study, kNN was adopted to find each user's nearest neighbors. Hence, the time complexity for finding all neighbors is $O(m\times n)$. Since the maximum number of ratings in the dataset is $m\times n$, all unrated items should be predicted by the proposed model, therefore, the overall time complexity for the whole dataset is $O(m^2n^2)$.

Table 5 provides a summary of the what the metric considers and what it ignores. In this study, the and the disadvantages of the metrics Adjusted Triangle and Jaccard hybridizations are designed to provide information on co-rated as well as non-co-rated users while the URP component was introduced to take into account the differences in the rating behavior of users.

Table 5. Summary of the merits and demerits of the proposed metrics.

Similarity measure	Considers	Ignores
Triangle	Both the length and angle between two vectors and the co-rated items	Items rated by either user and the user rating preference behavior
Adjusted Triangle	Both the length of the vectors and the angle between them	URP behavior
URP	the mean and deviation of the rating to model user preference	Need for more computations
Jaccard	Co-rated items	Absolute value of ratings

4. Experiments and Evaluation Metrics

4.1. Datasets Descriptions and Sparsity

4.1.1. Datasets

The effectiveness of the proposed similarity measure was validated through offline experiments performed on publicly available datasets. Offline evaluation is a popular and valuable means to investigate aspects of recommendation [28] because it does not require any interaction with real users and allows replicability and comparison of approaches at a low cost. We performed experiments on four datasets, that is, MovieLens 100k, MovieLens 1M, FilmTrust, and CiaoDVD, to evaluate the performance of the proposed similarity measurement method. Offline evaluations are a valuable means to investigate certain aspects of recommendation algorithms,

a) MovieLens

GroupLens Research collected and made available movie rating data from the MovieLens website. The MovieLens dataset [21] has been widely used in many scientific papers. The MovieLens 100k dataset consists of a collection of 100,000 movie ratings on a scale of 1-5 from 943 users' of 1682 movies. The MovieLens dataset 1m contains 1,000,209 anonymous movie ratings on a scale of 1 to 5 from 3,952 movies made by 6,040 users. In this dataset, each user rated at least twenty movies.

b) FilmTrust

The FilmTrust data were crawled from the FilmTrust website in June 2011 as part of a research paper on RS [19]. The FilmTrust dataset consists of 1508 users and 2,071 items. There was a total of 35,497 ratings, and the rating scale ranged from 0.5 4.

c) CiaoDVD

CiaoDVD is a DVD rating website where users share their movie reviews and provide recommendations for stores with the best prices. The dataset was crawled from a website as part of a research paper on trust prediction [20]. The dataset contains 17,615 users and 16,121 items. There were 72,664 ratings, and the rating scale ranged from 1 to 5.

4.1.2. Sparsity of the Datasets

In the matrix completion problem [27], the user-item rating matrix has missing data. The amount of missing data is called the sparsity level of the matrix. The sparsity level of the rating matrix is defined in [44] as follows:

$$Density = \frac{Non - Zero\ elements}{Total\ number\ of\ element} x100 = \frac{Total\ Ratings}{Users\ x\ Items} x100 \tag{14}$$

$$Sparsity = (100 - Density)\% \tag{15}$$

We selected these four datasets because they are the most commonly used datasets by researchers and industry in CF, CB, Hybrid, and trust-based RS. All the datasets were sparse. For instance, the density of the MovieLens 1M dataset is 1,000,209/6040×3952=4.19%; therefore, the sparsity level of the MovieLens 1M dataset is (100–4.19)=95.81%. Table 6 summarizes the datasets used in the experiments. All the datasets used were sparse, with CiaoDVD depicting the highest sparsity.

Table 6. Summary of datasets sparsity and density.

Dataset	No of Users	No of Items	No of Ratings	Sparsity (%)	Density (%)
ML-100K	943	1,682	100,000	93.70	6.30
ML-1M	6,040	3,952	1,000,209	95.81	4.19
FilmTrust	1,508	2,071	35,497	98.86	1.14
CiaoDVD	17,615	16,121	72,664	99.97	0.03

4.2. Evaluation Metrics

Recommendation accuracy and quality can be quantified using error metrics such as *MAE* and RMSE, or Accuracy metrics such as Precision, Recall and F-measure through offline studies of the algorithm's performance [52]. *MAE* measures the absolute value of the difference between the prediction (*p*) of the algorithm and the real rating (*r*). It is computed overall ratings available in the evaluation subset using Equation (16):

$$MAE = \frac{1}{N} \sum_{u,i} |p_{u,i-} r_{u,i}| \tag{16}$$

Where P_{ui} is the predicted rating for user u on item *i*, $r_{u,i}$ is the actual rating, and *N* is the total number of ratings for the item set. The *RMSE* is similar to *MAE*, but places more emphasis on larger deviations, given as

$$RMSE = \sqrt{\frac{1}{n} \sum_{u,i} (p_{u,i-} r_{u,i})^2} \tag{17}$$

MAE and *RMSE* [46] are widely used to evaluate the performance of recommendation systems. The lower the *MAE* and *RSME* values, the better the performance of the recommender system.

To measure the quality of the recommendations, we evaluated the *Precision*, *Recall* and the F-measure metrics. Measuring *Precision* and recall is required before calculating the F-measure, as the latter is the harmonic mean of *Precision* and *Recall*. When a recommendation system is tuned to increase *Precision*, *Recall* decreases and vice versa. Both metrics are inversely related, such that when *Precision* increases, Recall decreases. To balance the trade-offs between *Precision* and *Recall*, the F-measure is used, which is the weighted average of the *Precision* and *Recall*. The F-measure was calculated as the standard harmonic mean of *Precision* and *Recall* [23]. Therefore, we selected the F-measure as the evaluation metric, computed as shown in Equation (18):

$$Precision = \frac{n}{Top\ N} \tag{18}$$

where *n* is the number of items appearing in the recommended list, and is relevant to the testing user.

$$Recall = \frac{n}{M_T} \tag{19}$$

Where M_T is the total number of relevant items in the testing set.

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

The larger the *F-measure*, the higher the recommendation quality. We selected MAE, RMSE, and *F-measure* as evaluation metrics for the comparative study.

4.3. Baseline Methods

To elaborate on the efficiency of our proposed ATURPJ similarity measures, we compared it with the following baseline recommendation methods.

1. Triangle-multiply Jaccard. Sun *et al*. [47] proposed a

hybrid similarity measure of the integrated Triangle and Jaccard coefficient (TMJ) to improve recommendation accuracy.

2. Triangle multilevel. Alshammari *et al*. [4] proposed a Triangle multi-level item-based collaborative filtering approach with Triangle similarity measures that considers the length and angle of rating vectors between users, and allows positive and negative adjustments using a multi-level recommendation technique. The results of the Triangle multilevel similarity measure were compared with those of TMJ.
3. Factors impacting similarity. Feng *et al*. [17] proposed an improved similarity model that considers three impact factors of similarity to minimize the deviation of the similarity calculation:
   a) S1 which is used to define the similarity between users.
   b) S2 is introduced to punish user pairs with a small proportion of the number of co-rated items and
   c) S3 is adopted to weight each user's rating preference.

Feng *et al*. [17] used the proposed model for comparison with the TMJ. The proposed ATURPJ was compared with the proposed triangular multilevel and item-based TMJ.

4.4. Experiment Setup

The experiments were performed using an open-source Python library called SURPRISE, [22] which was used to build and analyze rating prediction algorithms. The SURPRISE library has built-in datasets, such as MovieLens and Jester, and includes common algorithms, including Cosine, Pearson and MSD similarity measures [30]. The implementation of the experiments involved programming the SURPRISE library to include additional similarity metrics such as Jaccard, Triangle, Adjusted Triangle, URP, TMJ, and the proposed similarity metric of ATURPJ. The SURPRISE library is licensed under the Creative Commons Attribution 4.0 International License.

We evaluated the performance of ATURPJ over four different datasets, MovieLens 100K, MovieLens 1M, FilmTrust, and CiaoDVD, and compared it with cosine, Pearson, Jaccard, Triangle and TMJ similarity metrics. TMJ is a product of the Triangle and Jaccard similarity measures. The performance of the experiments provides a comparison between ATURPJ, and the five other similarity measures on the four datasets described in section 4.1.1. The performance of the experiments followed a systematic pattern of importing the required libraries from the surprise library, loading the data, and performing an exploratory data analysis. This was followed by splitting the data into training and testing parts. The training set constituted 80% of the data, whereas the remaining 20% was the testing set. We then performed a 5-Fold cross-validation to obtain average results. This is followed by an evaluation before commencing the comparison of the similarity measure. The prediction accuracies of the MAE and RMSE are presented as the average of the five-fold results. The proposed similarity measure ATURPJ, was compared with Cosine, Pearson, Jaccard, Triangle, and TMJ. We begin with the MovieLens ML-100K dataset.

The experiments were conducted based on the item-based k-NN algorithm, and performance measures were calculated for different values of k. To examine the sensitivity of the neighborhood size and evaluate the quality of the recommendation, we performed additional experiments with different numbers of nearest neighbors and the recommended numbers N, and observed different experimental results.

4.5. Results and Discussion

In this section, the results, analysis, and comparison with baseline methods are presented.

4.5.1. MAE and RMSE Results

The average MAE and RMSE performance evaluation scores of the different similarity measures are shown in Table 7. From the results, it was observed that ATURPJ provided the best MAE and RMSE across all four datasets with different sparsity levels. The cosine similarity had the highest MAE and RMSE for the ML-100K dataset. TMJ showed good performance on the ML-100K dataset, coming fourth overall. Similarly, the Adjusted Triangle continuously outperformed the Triangle similarity across all the datasets marginally.

Table 7. Comparison of TURPJ performance with baseline methods.

Similarity metric	Dataset							
	ML-100K		ML-1M		Film trust		CiaoDVD	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Cosine	0.6994	0.8904	0.7018	0.8927	0.6258	0.8245	0.7883	0.9892
Person	0.6944	0.8857	0.6976	0.8889	0.6225	0.8302	0.7887	0.9971
Jaccard	0.796	0.8671	0.6809	0.8682	0.6208	0.822	0.7947	0.9934
Triangle	0.6916	0.8814	0.6938	0.8833	0.6251	0.823	0.7811	0.9895
TMJ	0.6775	0.8647	0.6786	0.8656	0.6201	0.8202	0.7944	0.9934
ATriangle	0.6829	0.8817	0.6843	0.8725	0.6225	0.8215	0.7825	0.9915
ATURPJ	0.5473	0.7347	0.5543	0.7554	0.5005	0.6927	0.5524	0.7216
TMJ [9]	0.707	0.903	0.671	0.859	0.614	0.816	-	-
Triangle multi-level [48]	0.77	1.05	0.83	1.05	-	-	-	-
Feng *et al*. method [17]	0.77	0.98	-	-	0.64	0.85	-	-

The highest MAE and RMSE values were obtained from the evaluation of the CiaoDVD datasets, which had the highest data sparsity. The comparison of our results with those of the baseline methods shows a significant difference with ATURPJ, which provides superior performance. In particular, the Adjusted Triangle result outperformed those of Sun *et al*. [47], Feng *et al*. [17], and Alshammari *et al*. [4] multilevel Triangle similarity measures.

a) MovieLens-100K dataset.

Figures 2 and 3 show the performance evaluation of the different similarity measures with varying numbers of k-nearest neighbors on the ML-100K dataset. The results show that ATURPJ is the best similarity measure, as it returns the lowest MAE and RMSE. The evaluated MAE and RMSE of all similarity measures decreased as the number of nearest neighbors increased. The performance of cosine is the worst, as it returns the highest MAE and RMSE, followed by Pearson coefficient similarity. The two plots clearly show that our proposed model surpasses the other similarity measures over the entire range of k.

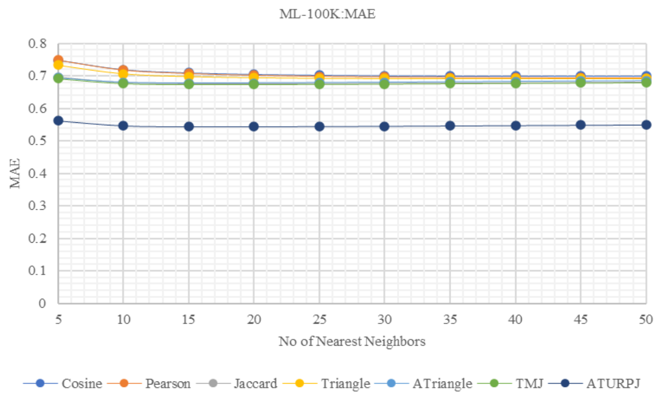

Figure 2. ML-100K MAE.

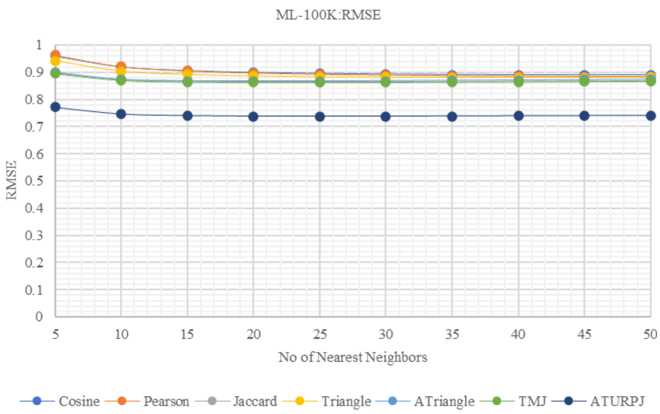

Figure 3. ML-100K: RMSE.

b) MovieLens-1M.

The performance evaluation of different similarity measures with varying numbers of k-nearest neighbors on the ML-1M dataset is shown in Figures 4 and 5. The proposed ATURPJ significantly outperformed all other similarity measures.

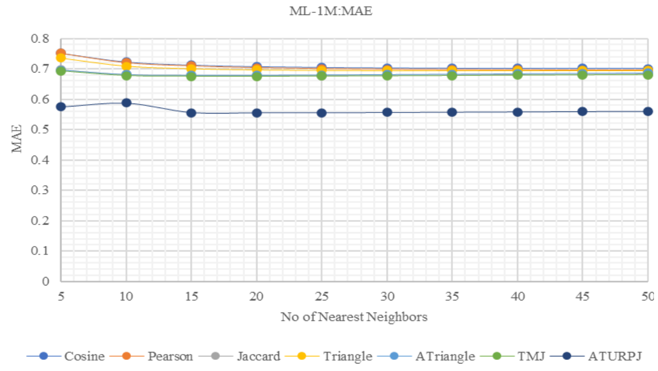

Figure 4. ML-1M MAE.

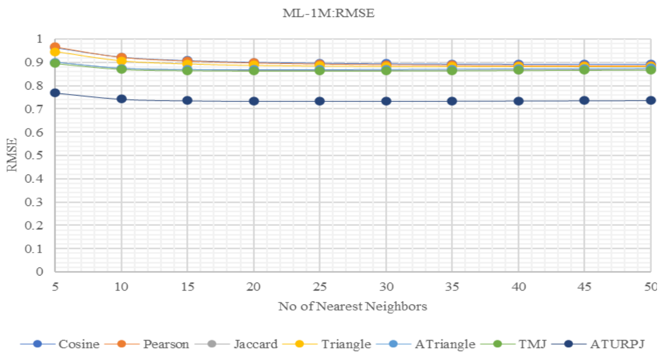

Figure 5. ML-1M: RMSE.

c) FilmTrust dataset.

For the FilmTrust dataset, the results are shown in Figures 6 and 7 for the MAE and RMSE. A slight drop in the MAE and RMSE was noted between k values of 5 and 10. Thereafter, the performance steadily decreased for all the similarity measures. The results show good performance of the ATURPJ.

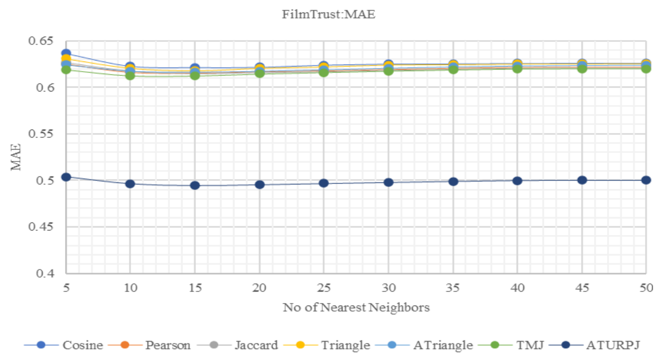

Figure 6. FilmTrust MAE.

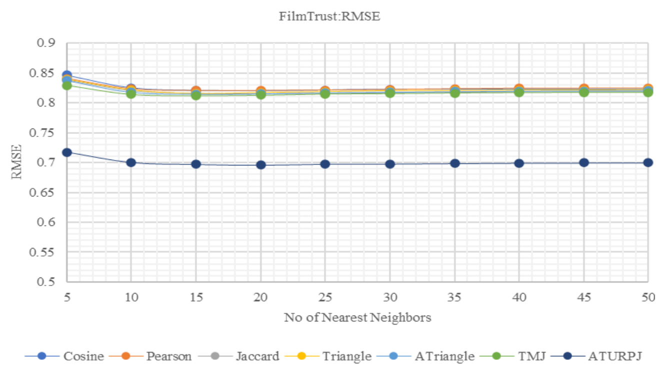

Figure 7. FilmTrust RMSE.

d) CiaoDVD dataset.

A similar performance trend for the MAE and RMSE is depicted for the CiaoDVD datasets, as shown in Figures 8 and 9. A slight drop was noted between k values of 5 and 10 for all similarity measures on the MAE and RMSE metrics. ATURPJ clearly outperformed all other existing measures. Jaccard and Pearson's performances were the lowest. ATURPJ provided a clear performance for the RMSE. As the value of k increased, the MAE and RMSE curves became constant.

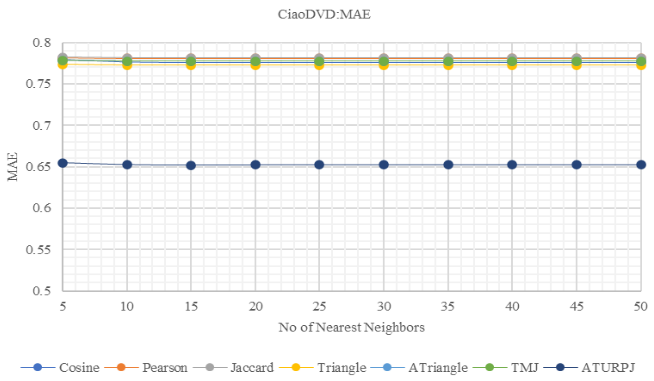

Figure 8. CiaoDVD MAE.

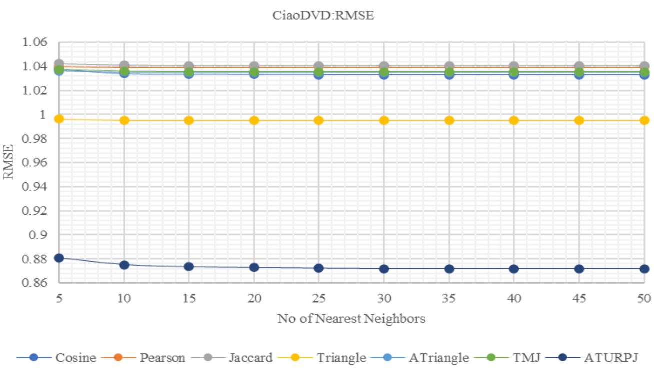

Figure 9. CiaoDVD RMSE.

4.5.2. F-Measure Results

The highest possible value of an F-measure is 1.0, indicating perfect Precision and recall, and the lowest possible value is 0 if either the Precision or Recall is zero. Table 8 gives a summary of the F-measures analysis for the various metrics across the datasets. ATURPJ has the highest F-measure score across all the datasets. Notably, an F-measure of 0.992 and 0.973 on the FilmTrust and CiaoDVD dataset which have a data sparsity of 98.86% and 99.97% respectively.

Table 8. F-measure analysis of the proposed hybrid measure against the existing metrics.

Similarity method	ML-100k	ML-1M	Film trust	CiaoDVD
Cosine	0.665	0.680	0.574	0.539
Person	0.645	0.618	0.679	0.641
Jaccard	0.719	0.718	0.676	0.740
Triangle	0.746	0.731	0.775	0.744
ATriangle	0.763	0.732	0.736	0.747
TMJ	0.749	0.739	0.786	0.744
ATURPJ	0.933	0.913	0.992	0.973

Figures 10, 11, 12, and 13 show the plots for the F-measure on the ML-100K, ML-1M, FilmTrust, and CiaoDVD datasets. ATURPJ had the highest F-measure. The performance of the cosine was the lowest in the ML-1ook and ML-1M datasets. All the similarity measures show a slight increase in the F-measure between k values of 5 and 10, with an upward trend thereafter, and then become steady. The performance remains almost constant after a k value of 25 and above. Similarly, an increase in the value of the F-measure occurs as the k value changes from 5 to 10 in all datasets, and for all similarity measures, the subsequent performance increases steadily thereafter. The trend of the curve became almost constant from the k value of 15.

Precision, Recall, and F-measure are key metrics that help in understanding how well a recommender system is performing. When applied at k, these metrics provide insights into the quality of the top k recommendations.

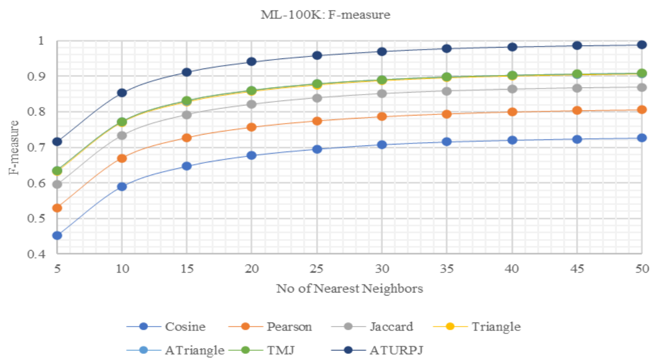

Figure 10. F-measure analysis of the proposed hybrid measure against the existing metrics on ML-100k dataset.

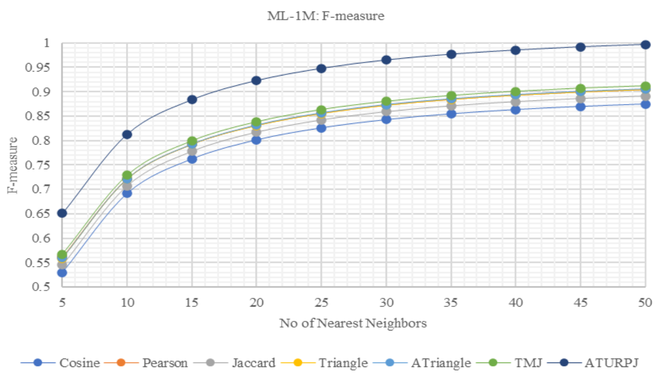

Figure 11. F-measure analysis of the proposed hybrid measure against the existing metrics on ML-1M dataset.

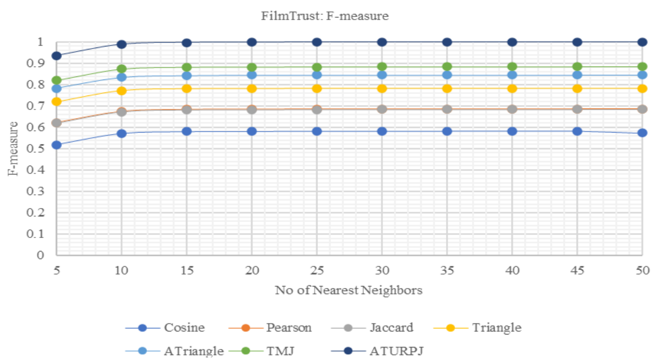

Figure 12. F-measure analysis of the proposed hybrid measure against the existing metrics on FilmTrust dataset.

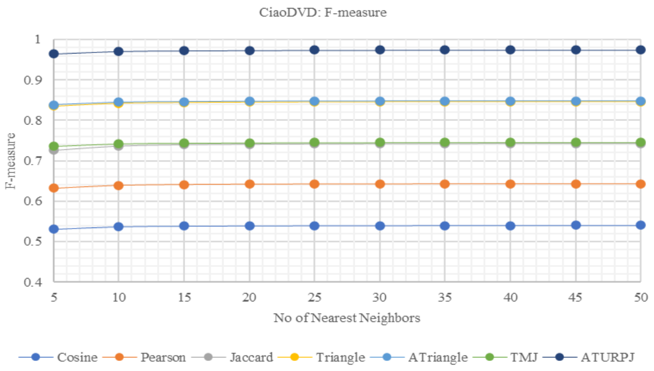

Figure 13. F-measure analysis of the proposed hybrid measure against the existing metrics on CiaoDVD dataset.

4.5.3. Significance of ATURPJ Results

To understand the significance of the results, we performed significance testing. Offline experiments are performed by computing the metrics repeatedly on the same historical datasets; thus, statistical significance testing can be performed offline using paired tests. Paired tests have the advantage of requiring fewer samples to determine a significant effect. Two-tailed Wilcoxon Signed-Rank tests at a significance level of 0.05 to ascertain whether the ATURP and ATURPJ similarity measure approach performed better than the conventional approaches. We paired all MAE results obtained from the similarity measure against the number of nearest neighbors. The statistical significance of an improvement is represented as a p-value using Student's t-test and the z-value using the Wilcoxon Signed-Rank test. When $p<\alpha(\alpha=0.05)$, the difference between the two systems was deemed to be statistically significant. The results of the two-tailed hypotheses show that all differences are statistically significant at $p<0.05$. The proposed hybrid ATURPJ similarity measures showed better performance than existing measures.

4.5.4. Discussion

The goal of this study is to enhance the prediction accuracy and improve the recommendation quality of collaborative filtering. In this paper, we presented an improved hybrid similarity measure based on Adjusted Triangles, the user rating preference and Jaccard coefficient. The proposed method was experimentally evaluated using four real-world datasets with varying degrees of sparsity to demonstrate the superior performance of the hybrid measures, and evaluated on well-known prediction accuracy and recommendation quality metrics. The different techniques in neighborhood-based approaches compute the similarity between users or items by representing user profiles as a vector of ratings given to individual products. The sparsity of the user-item rating matrix is an inherent problem in CF research and has a significant effect on the performance of a CF system. Because users often vote only on a tiny portion of the items included in the system, CF suffers from a significant degree of sparsity in the vote databases.

The proposed hybrid similarity measure showed better performance than the existing similarity metrics based on the results of the MAE, RMSE and the F-measure. Traditional similarity metrics, such as Pearson, Cosine, and MSD, were derived from statistics and are not appropriate in the field of RS because of extreme data sparsity and the limited range of acceptable vote values. When used in hybrid filtering, the RS can leverage clustering approaches to raise the prediction quality and lower the cold-start issue. The cold-start problem occurs when there are insufficient ratings at the beginning to allow for reliable recommendations. The fundamental tenet is that each user can receive appropriate recommendations based on the preferences of other similar users.

The results are significant as they clearly show that the similarity measure should take into consideration the URP as well as the commonly rated items and the items rated by either user in the recommendation matrix. The findings show that a different number of neighbors (K) will yield different prediction accuracy and quality of recommendation. The implications of these results are that the developed similarity metrics can easily be installed in existing neighborhood-based collaborative filtering, which has been extensively used to provide recommendations in various e-commerce applications in online business areas of travel, online broadcasting advertising, news, movies, and music.

5. Conclusions and Future Work

RS play an important role in e-commerce, health, news, scholarly articles and social networking domains in providing relevant information to users on items of interest [32, 51]. Similarity computation between users and items is a key component of any recommendation system. The performance of a recommendation system is highly dependent on the performance of the similarity-measurement stage. The existing similarity measure has drawbacks that limit its accuracy, as it does es not utilize all the ratings provided by the users. These setbacks were addressed using the CFs. Collaborative filtering techniques are implemented more frequently and often result in better predictive accuracy [34, 48]. These techniques recommend items based on the opinions of other like-minded users or identify items that are similar to those previously rated by the target user and mainly include item-based CF, which associates an item with nearest neighbors, and user-based CF, which associates a set of nearest neighbors with each user.

The study indicated that collaborative filtering techniques are used more regularly and frequently to achieve improved prediction accuracy. These techniques primarily use item-based CF, which associates an item with its nearest neighbors, and user-based CF, which associates a set of nearest neighbors with each user, to recommend items based on the ratings of other users who share their interests or find items that are similar to those previously rated by the target user. Slight gains in predictions and suggestions from a single type of information have been made by recent CF research (e.g., when the only information used is user ratings, information from social relations, or item content). When various CF algorithms and their associated data types are integrated, the results are further enhanced. The topic of hybrid CF techniques that employ current databases to simultaneously combine memory-based, social, and content-based information is gaining much attention from scholars. The hybrid similarity metric of ATURPJ will be a metric of choice in sparse datasets to give accurate predictions and quality recommendations.

From our assessment, it was observed that the ATURPJ measure consistently outperformed and produced better quality results than other similarity measures in terms of MAE, RMSE, and F-measure metrics. The results clearly show that the accuracy and quality of the recommendation are greatly improved and that the proposed hybrid similarity measures outperform the alternatives. Additionally, the recommendations are based on the ranking of how well the items within this set match the provided preferences. While accuracy metrics play a critical role in the evaluation of RS algorithms, there is need to explore non accuracy measures such as diversity, novelty, and serendipity, which is calibrated to user-specific preferences.

In future research, we would like to apply the improved hybrid similarity measure to context-aware recommendation systems.

References

[1] Abdalla H., Amer A., Amer Y., Nguyen L., and Al-Maqaleh B., "Boosting the Item-based Collaborative Filtering Model with Novel Similarity Measures," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, pp. 1-16, 2023. https://link.springer.com/article/10.1007/s44196-023-00299-2

[2] Al Hassanieh L., Abou Jaoudeh C., Bou Abdo J., and Demerjian J., "Similarity Measures for Collaborative Filtering Recommender Systems," *in Proceedings of the IEEE Middle East and North Africa Communications Conference*, Jounieh, pp. 1-5, 2018. DOI:10.1109/MENACOMM.2018.8371003

[3] Al-bashiri H., Abdulgabber M., Romli A., and Hujainah F., "Collaborative Filtering Similarity Measures: Revisiting," *in Proceedings of the International Conference on Advances in Image Processing*, Bangkok, pp. 195-200, 2017. https://doi.org/10.1145/3133264.3133299

[4] Alshammari G., Kapetanakis S., Polatidis N., and Petridis M., "A Triangle Multi-Level Item-Based Collaborative Filtering Method that Improves Recommendations," *in Proceedings of the 19th International Conference on Engineering Applications of Neural Networks*, Bristol, pp. 145-157, 2018. https://link.springer.com/chapter/10.1007/978-3-319-98204-5_12

[5] Amer A., Abdalla H., and Nguyen L., "Enhancing Recommendation Systems Performance Using Highly-Effective Similarity Measures," *Knowledge-Based Systems*, vol. 217, pp. 106842, 2021. https://doi.org/10.1016/j.knosys.2021.106842

[6] Arnold M., Goldschmitt M., and Rigotti T., "Dealing with Information Overload: A Comprehensive Review," *Frontiers in Psychology*, vol. 14, pp. 1-23, 2023. https://doi.org/10.3389/fpsyg.2023.1122200

[7] Ayub M., Ghazanfar M., Mehmood Z., Saba T., Alharbey R., Munshi A., and Alrige M., "Modeling User Rating Preference Behavior to Improve the Performance of the Collaborative Filtering Based Recommender Systems," *PLoS One*, vol. 14, no. 8, pp. 1-29, 2019. DOI:10.1371/journal.pone.0220129

[8] Bag S., Kumar S., and Tiwari M., "An Efficient Recommendation Generation Using Relevant Jaccard Similarity," *Information Sciences*, vol. 483, pp. 53-64, 2019. https://doi.org/10.1016/j.ins.2019.01.023

[9] Bobadilla J., Ortega F., and Hernando A., "A Collaborative Filtering Similarity Measure Based on Singularities," *Information Processing and Management*, vol. 48, no. 2, pp. 204-217, 2012. https://doi.org/10.1016/j.ipm.2011.03.007

[10] Bobadilla J., Ortega F., Hernando A., and Gutierrez A., "Recommender Systems Survey," *Knowledge-Based Systems*, vol. 46, pp. 109-132, 2013. DOI:10.1016/j.knosys.2013.03.012

[11] Breese J., Heckerman D., and Kadie C., "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, Wisconsin, pp. 43-52, 1998. https://dl.acm.org/doi/10.5555/2074094.2074100

[12] Burke R., "Hybrid Recommender Systems: Survey and Experiments," *User Modelling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370, 2002. https://link.springer.com/article/10.1023/A:1021240730564

[13] Chen R., Hua Q., Chang Y., Wang B., Zhang L., and Kong X., "A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks," *IEEE Access*, vol. 6, pp. 64301-64320, 2018.

https://ieeexplore.ieee.org/document/8506344
[14] Chen Z., Gan W., Wu J., Hu K., and Lin H., "Data Scarcity in Recommendation Systems: A Survey," *ACM Transactions on Recommender Systems*, vol. 1, no. 1, pp. 1-32, 2024. https://doi.org/10.1145/3639063
[15] Donchev D., 37 Mind Blowing YouTube Facts, Figures and Statistics, Fortunelords, https://fortunelords.com/youtube-statistics/, Last Visited, 2024.
[16] Fayyaz Z., Ebrahimian M., Nawara D., Ibrahim A., and Kashef R., "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities," *Applied Sciences*, vol. 10, no. 21, pp. 1-20, 2020. DOI:10.3390/app10217748
[17] Feng J., Feng X., Zhang N., and Peng J., "An Improved Collaborative Filtering Method Based on Similarity," *PLoS One*, vol. 13, no. 9, pp. 1-18, 2018. DOI:10.1371/journal.pone.0204003
[18] Fkih F., "Similarity Measures for Collaborative Filtering-based Recommender Systems: Review and Experimental Comparison," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 7645-7669, 2022. https://doi.org/10.1016/j.jksuci.2021.09.014
[19] Guo G., Zhang J., and Yorke-Smith N., "A Novel Evidence-Based Bayesian Similarity Measure for Recommender Systems," *ACM Transactions on the Web*, vol. 10, no. 2, pp. 1-30, 2016. https://doi.org/10.1145/2856037
[20] Guo G., Zhang J., Thalmann D., and Yorke-Smith N., "ETAF: An Extended Trust Antecedents Framework for Trust Prediction," *in Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Beijing, pp. 540-547, 2014. DOI:10.1109/ASONAM.2014.6921639
[21] Harper F. and Konstan J., "The MovieLens Datasets: History and Context," *ACM Transactions on Interactive Intelligent Systems*, vol. 5, no. 4, pp. 1-19, 2016. https://doi.org/10.1145/2827872
[22] Hug N., "Surprise: A Python Library for Recommender Systems," *Journal of Open-Source Software*, vol. 5, no. 52, pp. 2174, 2020. DOI:10.21105/joss.02174
[23] Isinkaye F., Folajimi Y., and Ojokoh B., "Recommendation Systems: Principles, Methods and Evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261-273, 2015. https://doi.org/10.1016/j.eij.2015.06.005
[24] Jaccard P., "The Distribution of the Flora in the Alpine Zone," *New Phytologist*, vol. 11, no. 2, pp. 37-50, 1912. https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1912.tb05611.x
[25] Jain G., Mahara T., and Tripathi K., *Soft Computing: Theories and Applications*, Springer, 2020. https://link.springer.com/chapter/10.1007/978-981-15-0751-9_32
[26] Jain G., Mahara T., Sharma S., and Sangaiah A., "A Cognitive Similarity-Based Measure to Enhance the Performance of Collaborative Filtering-based Recommendation System," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 6, pp. 1785-1793, 2022. https://ieeexplore.ieee.org/document/9829416
[27] Jannach D. and Zanker M., *Collaborative Recommendations*, World Scientific, 2018. https://doi.org/10.1142/9789813275355_0001
[28] Jannach D., Pu P., Ricci F., and Zanker M., "Recommender Systems: Past, Present, Future," *Association for the Advancement of Artificial Intelligence Magazine*, vol. 42, no. 3, pp. 1-6, 2021. https://doi.org/10.1609/aimag.v42i3.18139
[29] Kaminskas M. and Bridge D., "Diversity, Serendipity, Novelty, and Coverage," *ACM Transactions on Interactive Intelligent Systems*, vol. 7, no. 1, pp. 1-42, 2017. DOI:10.1145/2926720
[30] Kamta S. and Verma V., *Evolutionary Computing and Mobile Sustainable Networks*, Springer, 2021. https://doi.org/10.1007/978-981-15-5258-8_80
[31] Kapoor K., Tamilmani K., Rana N., Patil P., Dwivedi Y., and Nerur S., "Advances in Social Media Research: Past, Present and Future," *Information Systems Frontiers*, vol. 20, no. 3, pp. 531-558, 2018. https://link.springer.com/article/10.1007/s10796-017-9810-y
[32] Knudsen E., "Modeling News Recommender Systems' Conditional Effects on Selective Exposure: Evidence from two Online Experiments," *Journal of Communication*, vol. 73, no. 2, pp. 138-149, 2023. https://doi.org/10.1093/joc/jqac047
[33] Koutrika G., Bercovitz B., and Garcia-Molina H., "FlexRecs: Expressing and Combining Flexible Recommendations," *in Proceedings of the International Conference on Management of Data and 28th Symposium on Principles of Database Systems*, Rhode Island, pp. 745-757, 2009. https://doi.org/10.1145/1559845.1559923
[34] Liu H., Hu Z., Mian A., Tian H., and Zhu X., "A New User Similarity Model to Improve the Accuracy of Collaborative Filtering," *Knowledge-Based Systems*, vol. 56, pp. 156-166, 2014. https://doi.org/10.1016/j.knosys.2013.11.006
[35] Manochandar S. and Punniyamoorthy M., "A New User Similarity Measure in a New Prediction Model for Collaborative Filtering," *Applied Intelligence*, vol. 51, no. 1, pp. 586-615, 2021. https://doi.org/10.1007/s10489-020-01811-3
[36] Manouselis N., Drachsler H., Verbert K., and Santos O., "Workshop on Recommender Systems

for Technology Enhanced Learning," *in Proceedings of the 4th ACM Conference on Recommender Systems*, Barcelona, pp. 377-378, 2010. https://dl.acm.org/doi/10.1145/1864708.1864797

[37] Marsh E., Perez Vallejos E., and Spence A., "Overloaded by Information or Worried about Missing out on it: A Quantitative Study of Stress, Burnout, and Mental Health Implications in the Digital Workplace," *Sage Open*, vol. 14, no. 3, pp. 1-17, 2024. https://doi.org/10.1177/21582440241268830

[38] Meissa M., Benharzallah S., Kahloul L., and Kazar O., "A Personalized Recommendation for Web API Discovery in Social Web of Things," *The International Arab Journal of Information Technology*, vol. 18, no. 3A, pp. 438-445, 2021. https://doi.org/10.34028/iajit/18/3A/7

[39] Nijor S., Rallis G., Lad N., and Gokcen E., "Patient Safety Issues from Information Overload in Electronic Medical Records Background and Objective: Electronic Health Records (EHRs) Have," *Journal of Patient Safety*, vol. 18, no. 6, pp. 999-1003, 2022. https://pmc.ncbi.nlm.nih.gov/articles/PMC9422765/

[40] Pearson K., "Note on Regression and Inheritance in the Case of two Parents," *Proceedings of the Royal Society of London*, vol. 58, no. 347-352, pp. 240-242, 1895. https://royalsocietypublishing.org/doi/abs/10.1098/rspl.1895.0041

[41] Resnick P., Iacovou N., Suchak M., Bergstrom P., and Riedl J., "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," *in Proceedings of the ACM Conference on Computer Supported Cooperative Work*, Carolina, pp. 175-186, 1994. https://doi.org/10.1145/192844.192905

[42] Roetzel P., "Information Overload in the Information Age: A Review of the Literature from Business Administration, Business Psychology, and Related Disciplines with a Bibliometric Approach and Framework Development," *Business Research*, vol. 12, no. 2, pp. 479-522, 2019, https://link.springer.com/article/10.1007/s40685-018-0069-z

[43] Roy D. and Dutta M., "A Systematic Review and Research Perspective on Recommender Systems," *Journal of Big Data*, vol. 9, no. 1, pp. 1-36, 2022. https://doi.org/10.1186/s40537-022-00592-5

[44] Sarwar B., Karypis G., Konstan J., and Riedl J., "Analysis of Recommendation Algorithms for e-Commerce," *in Proceedings of the 2nd ACM Conference on Electronic Commerce*, Minnesota, pp. 158-167, 2000. https://dl.acm.org/doi/10.1145/352871.352887

[45] Shi Y., Larson M., and Hanjalic A., "Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges," *ACM Computing Surveys*, vol. 47, no. 1, pp. 1-45, 2014. https://dl.acm.org/doi/10.1145/2556270

[46] Silveira T., Zhang M., Lin X., Liu Y., and Ma S., "How Good your Recommender System is? A Survey on Evaluations in Recommendation," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 5, pp. 813-831, 2019. https://link.springer.com/article/10.1007/s13042-017-0762-9

[47] Sun S., Zhang Z., Dong X., Zhang H., Li T., Zhang L., and Min F., "Integrating Triangle and Jaccard Similarities for Recommendation," *PLoS One*, vol. 12, no. 8, pp. 1-16, 2017. https://doi.org/10.1371/journal.pone.0183570

[48] Suryakant and Mahara T., "A New Similarity Measure Based on Mean Measure of Divergence for Collaborative Filtering in Sparse Environment," *Procedia Computer Science*, vol. 89, pp. 450-456, 2016. https://doi.org/10.1016/j.procs.2016.06.099

[49] Yuan J., Shalaby W., Korayem M., Lin D., Aljadda K., and Luo J., "Solving Cold-Start Problem in Large-Scale Recommendation Engines: A Deep Learning Approach," *in Proceedings of the IEEE International Conference on Big Data, Big Data*, Washington (DC), pp. 1901-1910, 2016. DOI:10.1109/BIGDATA.2016.7840810

[50] Zhang B. and Yuan B., "Improved Collaborative Filtering Recommendation Algorithm of Similarity Measure," *AIP Conference Proceedings*, vol. 1839, no. 1, 2017. https://doi.org/10.1063/1.4982532

[51] Zhang Z., Patra B., Yaseen A., Zhu J., Sabharwal R., Roberts K., Cao T., and Wu H., "Scholarly Recommendation Systems: A Literature Survey," *Knowledge and Information Systems*, vol. 65, no. 11, pp. 4433-4478, 2023. DOI:10.1007/s10115-023-01901-x

[52] Zolaktaf Z., AlOmeir O., and Pottinger R., "Bridging the Gap Between User-Centric and Offline Evaluation of Personalized Recommendation Systems," *in Proceedings of the Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*, Singapore, pp. 183-186, 2018. DOI:10.1145/3213586.3226216

Mohamed Abdi is a Ph.D. student in Information Technology at the Jomo Kenyatta University of Agriculture and Technology (JKUAT), Kenya. Mr. Abdi holds a Bachelor of Science Degree from Egerton University, Njoro, Kenya and a Master of Science (Computer Applications) Degree from Hohai University, Nanjing, China. He is also a Microsoft Certified Systems Engineer (MCSE). His research interests are in Machine Learning, Recommender Systems, Data analytics, e-Commerce, e-Government, Security, Privacy and Trust.

George Okeyo is an Assistant Teaching Professor at Carnegie Mellon University Africa, Kigali, Rwanda. He is an experienced Academic in Computer Science with a demonstrated history of working in the higher education industry as Lecturer, Researcher, Mentor, Manager and Academic Head of Department. Skilled in Artificial Intelligence, Data Analytics, Web Development, and Programming and Software Development in Java, C/C++, PHP and Python. He holds Ph.D. Degree in Philosophy in Computer Science from University of Ulster. Prof. Okeyo is a recipient of the 2024 Carnegie Mellon University's College of Engineering Dean's Early Career Fellowship.

Ronald Mwangi is an Associate Professor of Computing at Jomo Kenyatta University of Agriculture and Technology (JKUAT), School of computing and Information Technology, Kenya. Bachelor degree in Mathematics at Kenyatta University (Kenya) in 1989, then MSc in Operation Research in Shanghai University (China) in 1995, followed by PhD in Systems and Information Engineering in Hokkaido University (Japan) in 2004. His research areas include; Systems Modelling and Development, Smart Agent Computing, Simulation and ICT Policy Formulation.