Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

Covering Rough Set based Collaborative Filtering Technique: Application for Social Tag Recommender Systems

Dr Mohamed Doheir
Senior lecturer, Faculty of Technology Management & Technopreneurship,
Universiti Teknikal Malaysia
doheir@utem.edu.my

Abstract:

Recommendation systems have become an integral part of our daily lives, providing personalized information and content tailored to individual preferences. Collaborative filtering (CF) is a widely adopted technique in recommendation systems, excelling at delivering high-quality recommendations by identifying users with similar preferences based on their past interactions and history. The Covering Rough Set (CRS) model introduces a unique approach where relevant items from each user's neighborhood collectively form a common covering. These common coverings, in turn, construct a covering set for an active user within a specific sphere. Employing covering reduction techniques helps eliminate redundant common coverings, optimizing the recommendation process. In this paper, we present a novel approach, the "Covering Rough Set-Based Collaborative Filtering Technique" (CRS-CF). This technique empowers users by learning weights on various features and harnesses rough set theory for the efficient representation of user characteristics. CRS-CF offers a personalized and robust recommendation mechanism by combining the strengths of CF and covering-based rough sets. Our study demonstrates the effectiveness of the proposed CRS-CF approach through comprehensive experimental results. We evaluate its performance using various criteria and a Social Tagging dataset. The results underscore the superiority of our approach in providing accurate and tailored recommendations, reaffirming the potential of the CRS-CF model in enhancing recommendation systems and furthering the field of personalized content delivery.

Keywords: Covering Rough Set, Collaborative Filtering, Social Tagging Systems, Recommender Systems, Root Mean Squared Error, and Mean Absolute Error.

1. Introduction

With the development of the internet and Computational Intelligent Techniques, the recommender system (RS) has become very popular recently. The RS can advise users when making decisions on the basis of personal preferences and help users discover items they might not find by themselves. RS use knowledge discovery and statistical methods for recommending items to users [1]. In any RS that uses collaborative filtering methods, computation of similarity metrics is a primary step to find out similar users or items.

Social Tagging Systems (STS) and data mining have gained interest among researchers and practitioners in the recent past all over the universe. Tags allow users to effectively annotate resources using keywords to personalize their recommendations and organize the resources for easy recovery. The STS is an application of social media that has succeeded as a substance to ease information search and sharing. In STS, Tagging can be regarded as the act of linking of entities such as users, resources and tags. It helps user better way to understand and disseminate their collections of attractive objects. When a user employs a tag to a resource in the system, a multilateral relationship between the user, the resource and the tag is formed. The tag recommendation system becomes useful by suggesting a lot of relevant keywords to annotate the resources.

CF techniques make recommendations for a given user by collecting data about users having similar tastes for the dedicated user. Thus, collaborative recommender systems allow personalized information tailored to the individual user's preferences [2]. The underlying assumption of CF is that if users have similar tastes (e.g., rating, buying, seeing, listening) then they will rot or act on other items similarly.

Tagging is a process in which a user can give meaningful terms to a resource to facilitate the easy discoverability of the resource. Tags are the nonhierarchical keywords of a resource, i.e., bookmarking, picture, or file [2]. Tagging allows the user to categorize the web resources, such as web pages, blog spots, pictures, multimedia images, and so on based on their content. Thus, the main objective of the tagging system is to structure and manage the web content and to discover the relevant content shared by other users. In Web 2.0 applications, a large number of tagging systems are available, e.g., Delicious, Flickr, BibSonomy and so forth.

The main purpose of tagging is to categorize the web resources based on their content. If many users apply the same word to tag an item, the tag will become great and sheer[11]. Tag recommendation supports a user to post his/her blog by recommending latent-related tags. Recommendation process is a greatest investigated scenario in folksonomy context.

Rough set theory was first presented by Pawlak in the early 1980s [3]. Covering-based rough set (CRS) has been regarded as a meaningful extension of the classical rough set to handle vague and imperfect knowledge better, which extends the partition of rough set to a covering [4,5]. Currently, much of the literature has been focused on providing the theory behind covering-based rough set [6,7], but there is little regarding applications, especially for RSs [8].

In this paper, a new model is introduced using CRS theory used for Collaborative Filtering of social tagging systems and the technique is implemented and examined using Social Tagging dataset. The Proposed Techniques are used to improve the CF approach. We obtain predictions and recommendations to attain more efficiency using CRS-CF. At long last this calculation is contrasted and existing approach Classical CF is compared by utilizing the measurements Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).

The proposed work consists of four major tasks:

1. Data Extraction: Fetching data from the Social Tagging Systems.
2. Data Formatting: Data formatting which consists of mapping the tags and users based on tag weights Represented in matrix format
3. Recommendation: To Recommend Tags to the Users Using Proposed Approach.
4. Pattern Analysis: MAE and RMSE are used to find the Accuracy.

The rest of this paper is organized as follows: Section 2 presents some of the related work. Section 3 Present Methodology of this research work. In Section 4, the experimental results have been reported. And the conclusion has been addressed in Section 5.

2. Related work

This section gives a brief review about Collaborative Filtering Techniques. The objective of our research work is to create and assess another direction, i.e Personalized Tag Recommendation using Collaborative filtering Techniques, for future web data mining. Collaborative Filtering with optimization technique have been extensively studied by some researchers. Collaborative filtering (CF) is a significant component of the recommendation process that is based on the ways in which humans have made decisions throughout history. Rough set theory is a mathematical tool to deal with vagueness and uncertainty of imprecise data.

The theory introduced by Pawlak has been developed and found applications in the fields of decision analysis, data analysis, pattern recognition, machine learning, expert systems, and knowledge discovery in databases, among others. This paper discusses covering rough set as the main research tool [10]. Covering-based rough set theory is a generalization of rough set theory. However, these measurements in rough sets cannot be used in covering based rough sets. Therefore, there is much need to construct some measurements in covering-based rough sets [9]. Furthermore, the structure of covering-based rough sets for recommendation have been a hotspot of study.Covering-based rough set has been regarded as a meaningful extension of the classical rough set to handle vague and imperfect knowledge better, which extends the partition of rough set to a covering. The notion of reduction for covering is one of the most important results in covering-based rough set [11]. Currently, much of the literature has been focused on providing the theory behind covering-based rough set, but there is little regarding applications, especially for RSs.

The collaborative filtering system could automatically filter the information that the system could not analyze and represent, and recommend up-to-date information. Collaborative filtering methods are based on collecting and analyzing a large amount of information based on users' behavior, activity or preferences and predicting what users will like based on their similarity to other users [12]. For user-based CF, if fewer users at the top of a similar list are selected as neighbors of an active user, high-accuracy items could be recommended for the active user; however, the types of recommendations will be decreased, even in just making the most popular items as recommendations.

If more types of items are to be recommended, more users should be selected as neighbors of the active user, but the accuracy will decrease as the number of neighbors grow. Therefore, it is difficult for CF to simultaneously obtain good values for metrics of accuracy and coverage [13]. To solve this problem, the relative effective neighbors should be selected from all neighbors such that the recommendations not only maintain good values of accuracy but also obtain satisfactory values of coverage. We observe here the definition of covering is an annex of the definition of partitions. Different lower and upper approximation operations would generate different types of covering-based rough set. The covering-based rough set was first presented by Zakowski [14], who extended

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

Pawlak's rough set theory from a division to a crossing. Pomykala gave the feeling of the second case of covering based rough set, while Tsang presented the third type [15], Zhu defined the fourth and fifth types of covering-based rough set models, and Wang studied the sixth type of covering-based approximations [16].

3. Proposed Methodology

The Proposed Methodology for Tag Recommendation comprises the following steps:

a. Data extraction
b. Data formatting
c. Recommendation
d. Pattern Analysis

1) Data Extraction

The experimental dataset can be extracted from the social tagging system, which helps users to search relevant resources using tags. The Social tagging system functions as a facilitator, enabling users to navigate and retrieve materials that are pertinent to their interests. Tags, which are descriptive keywords or labels applied to various items within the system, serve as a crucial mechanism for users to pinpoint and access the resources they seek. This integration of tags enhances the efficiency of resource discovery, making it more user-friendly and effective. Table 1 shows the example dataset extracted from the social tagging system.

Table 1: Social Tagging dataset

User id	Tag name	Tag count
1	Android	7
2	Java	2
1	Networking	8
1	Distributed system	9
2	Operating system	4
3	Grid computing	9
2	Cloud computing	5
4	java	7

2) Data Formatting

After fetching the dataset from the Social Tagging Data, the next step is to format the data set. i.e converting the dataset into matrix representation. Initially, the data is fetched from this source, which typically includes a wealth of information tagged by users. Following this data retrieval, the subsequent step involves data formatting. This entails a critical transformation of the dataset into a matrix representation. In the context of data analysis and machine learning, representing the data as a matrix is pivotal, as it enables various computational and analytical techniques. This matrix representation simplifies data manipulation and allows for the application of algorithms

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

that can uncover patterns, associations, and insights within the dataset, making it a crucial preparatory step in data analysis. Table 2 shows the matrix representation of the tag dataset. Rows corresponds to tags and columns corresponds to the users.

In the tag matrix,

- n represents tags, M represents users, Wij represents tag weight associated with users.

Table 2: Dataset Format

	User1	User2	User3	User4	User 5
Android	7	9	18	6	5
java	10	14	11	2	2
Networking	0	6	9	41	9
Distributed system	0	14	23	30	7
Operating system	0	1	19	4	8
Grid computing	5	0	25	3	2

3) Proposed Approach (CRS-PSO Collaborative Filtering)

Collaborative filtering RS

The Collaborative Filtering has turned into the most broadly utilized strategy to prescribe tags for users. In the realm of recommendation systems, collaborative filtering stands as a foundational approach, particularly crucial in the context of social tagging systems. These systems rely on user-generated tags to describe and categorize content, be it articles, images, or products. Collaborative filtering, within this context, strives to deliver personalized recommendations by analyzing user interactions, discerning patterns, and exploiting tag-based metadata. The Collaborative Filtering incorporates memory-based technique and model-based scheme. The memory-based system first computes the similitude among users and chooses the most comparable users as the neighbors. Recommendation model helps users to find out their potential future likes and interests. It recommends good products to users and satisfies the users' demands as far as possible.

Covering Rough Set-Based Recommendation System stands out as an intelligent and adaptable approach in the realm of recommendation systems. It adeptly handles uncertainty and incompleteness in data while delivering highly personalized and efficient recommendations. The convergence of Covering Rough Set Theory (CRST) and Particle Swarm Optimization (PSO) has given rise to an innovative approach in recommendation systems, known as CRST-PSO based recommendation systems. This hybrid methodology incorporates the strengths of both CRST and PSO to enhance the precision, personalization, and efficiency of recommendation algorithms.

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

Formally, in CF we have a set of users U = {u_1, u_2, . . . , u_p} and a set of items I = {i_1, i_2, . . . , i_q} such as songs, books, news articles, or movies. Ratings are stored in a p × q user-item rating matrix.

Definition 1. Let U be the domain of discourse and C be a family of subsets of U. If none of the subsets in C is empty, and ∪C = U, C is called a covering of U.

Definition 2. Let U be a non-empty set and C be a covering of U. We call the ordered pair <U,C> a covering approximation space.

Definition 3. Let C be a covering of a domain U and K ∈ C. If there exists another element K' of C such that K ⊂ K', we say that K is reducible in C;Otherwise, K is irreducible. When we remove all reducible elements from C, the new irreducible covering is called reduct of C and denoted by reduct(C).

Model Constructions

In this subsection, we present detailed information and the steps comprising CRS-CF, which does not use any user demographic data. In short, CRS-CF needs the following information:

The users set U: U = {u_1, u_2...u_E}, where E is the number of users.The items set S: S = {s_1, s_2...s_I}, where I is the number of items or tag.

The item's attributes set A: A = {α_1, α_2...α_P }, where α_n is an attribute of the item and P is the number of attributes.

Algorithm : CRS-Collaborative filtering
Input: — Social Tagging Data Set Output: — A set of recommended items Rec ⊂ S Step 1: Set Rec ={ ∅}. Step 2: Construct indiscernibility Relation IND(B) and Complementary sets of Tagged Bookmarks using Eq .1 $[x]B = \cap\{[(a, v)] \mid a \varepsilon B, \rho(x,a)= v$ Step 3: The Resultant Complementary Set is assigned to recommendation task. If two or more objects came in single set perform union operation Step 4: Find Cosine-based similarity approach to compute the similarity between the tags. $sim\,(a,b) = \frac{\sum_{u\in U_a\cap U_b}(r_{u_i,a}-\overline{r_a})(r_{u_i,b}-\overline{r_b})}{\sqrt{\sum_{u\in U_a\cap U_b}(r_{u_i,a}-\overline{r_a})^2}\ \sqrt{\sum_{u\in U_a\cap U_b}(r_{u_i,b}-\overline{r_b})^2}}$ Where $sim(a,b)$ indicates the similarity between tag a and b and S_{ab} is the set of all items rated by both tag a and b. Step 5 : Select Neighbor If Rsim (a,b) > γ , Then b is the neighbor of a Step 6: Predict and recommend the Tag for User by $P_{u,a} = \overline{r_a} + \frac{\sum_{b\in N(a)}(r_{u,b}-\overline{r_b}) * R_{sim}(a,b)}{\sum_{b\in N(a)} R_{sim}(a,b)}$ In the above, $P_{u,a}$ is the prediction of item a for user u. Step 7: Set Rec = D_N; output Rec.

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

We provide an example of Social tag data for centroid initialization using Covering Rough Set

Initialization:

Let t= {t1, t2, t3, t4} be the set of tags and b= {bm1, bm2, bm3} be the set of distinct bookmarks. Let t1 ={bm1,bm2} , t2={bm2,bm3},t3={bm1,bm3},t4={bm2,bm3}.

	BM1	BM2	BM3
T1	1	1	0
T2	0	1	1
T3	1	0	1
T4	0	1	1

Then the tag can be represented as vectors.
t1={1,1,0},t2={0,1,1},t3={1,0,1},t4={0,1,1}

Step1: Indiscernibility Relation and Complementary sets

In Table 1, the set of cases U = {T1, T2, T3, T4} and the set of attributes A = {BM1, BM2, BM3}. Rough set theory is based on the idea of an indiscernibility relation, defined for complete decision tables. The indiscernibility relation IND (B) may be computed using the idea of blocks of attribute-value pairs where B is a nonempty subset of the set A of all attributes.

[(BM1, 1)] = {1, 3},
[(BM1, 0)] = {2, 4},
[(BM2, 1)] = {1, 2, 4},
[(BM2, 0)] = {3},
[(BM3, 1)] = {2, 3, 4},
[(BM3, 0)] = {1}

The indiscernibility relation IND (B) is known when all elementary blocks of IND (B) are known. Such elementary blocks of B are intersections of the corresponding attribute-value pairs, i.e., for any case x ÎU,

Compute Elementary Sets

[(BM1, 1)] ∩ [(BM2, 1)] ∩ [(BM3, 1)]={1,3},{1,2,4},{2,3,4}={ϕ}
[(BM1, 1)] ∩ [(BM2, 1)] ∩ [(BM3, 0)] = {1, 3}, {1, 2, 4}, {1} = {1}
[(BM1, 1)] ∩ [(BM2, 0)] ∩ [(BM3, 1)] = {1, 3}, {3}, {2, 3, 4} = {3}
[(BM1, 1)] ∩ [(BM2, 0)] ∩ [(BM3, 0)] = {1, 3}, {3}, {1} = {ϕ}
[(BM1, 0)] ∩ [(BM2, 1)] ∩ [(BM3, 1)]= {2,4},{1,2,4},{2,3,4}={2,4}
[(BM1, 0)] ∩ [(BM2, 1)] ∩ [(BM3, 0)] = {2, 4}, {1, 2, 4}, {1} = {ϕ}
[(BM1, 0)] ∩ [(BM2, 0)] ∩ [(BM3, 1)] = {2, 4}, {3}, {2, 3, 4} = {ϕ}
[(BM1, 0)] ∩ [(BM2, 0)] ∩ [(BM3, 0)] = {2, 4}, {3}, {1} = {ϕ}

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

Step 2: The Resultant Complementary Set is assigned to recommendation task. If two or more objects came in single set perform union operation. Then replace the original values

	BM1	BM2	BM3
T1	4	2	0
T2,4	0	5	1
T3	3	0	2

Step 3: Similarity computation: Item similarity computation is the technique for computing similarity between the tags. This can be done by identifying the users those who are given weights for similar tags. Based on this, the similarity computation techniques are applied to determine the similarity between the tags. Let i and j be the tags. Then the similarity between the tags are represented as $sim_{i,j}$. In this work, correlation based similarity computation technique is adapted. In this case the similarity is computed based on the correlation among the users. Pearson correlation coefficient is the preferred choice. Let u_i are the users who are given weights both the tags a and b. then the similarity between the tags is computed as follows.

$$sim\,(a,b) = \frac{\sum_{u\in U_a\cap U_b}\left(r_{u_i,a}-\overline{r_a}\right)\left(r_{u_i,b}-\overline{r_b}\right)}{\sqrt{\sum_{u\in U_a\cap U_b}\left(r_{u_i,a}-\overline{r_a}\right)^2}\ \sqrt{\sum_{u\in U_a\cap U_b}\left(r_{u_i,b}-\overline{r_b}\right)^2}}$$

Here u_a is a set of users who rated a while u_b is a set of users who rated b, u is a user who both rated a and b, $r_{u,a}$ is the rating of a given by u, $r_{u,b}$ is the rating of b given by u and r_a is the average rating of a , and r_b is the average rating of b. The value of sim(a, b) is in the interval of [-1,1].

	Recovery tag pairs	Rating similarity
Rating similarity with respect to Tag 2	(Tag 2, Tag 1)	0.438
	(Tag 2, Tag 3)	0.10

Step 4: Neighborhood selection: The next step in recommendation algorithm is neighborhood selection. The neighbors for the recovery tag is selected. This is done by comparing the similarity value with the threshold value. If the similarity value exceeds threshold value, then that tag is considered as neighbor for the recovery tag. The neighbors of the target tag are determined according to the following formula

$$Neighbor(a) = \{b \mid R_{sim}(a,b) > \gamma, a \neq b\}$$

Here $R_{sim\ (a,b)}$ is the rating similarity between tag a and tag b. γ is the rating similarity threshold. The neighborhood Selection for the target recovery tag is calculated, the rating similarity threshold value is set as γ =0.4. (Tag 2, Tag 1) pair is positive and it is greater than the threshold value, hence Tag 1 is chosen as the neighbour of Tag 2. i.e.

Neighbour (Tag 2) =Tag 1

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

Step 5: Prediction computation: Based on the predicted rate, the tags are recommended to the users. Let u be the active user and a be the recovery tag, then the predicting rate p $_{(u,a)}$ is computed as follows

$$P_{u,a} = \overline{r_a} + \frac{\sum_{b \in N(a)} (r_{u,b} - \overline{r_b}) * R_{sim}(a,b)}{\sum_{b \in N(a)} R_{sim}(a,b)}$$

$\overline{r_a}$ is the average rate given to tag a. $b \in N(a)$ is the neighbor set of tag a. $r_{u,b}$ is the rate given by the user u to tag b. $R_{sim}(a,b)$ is the rating similarity between tag a and tag b. Using above Equation to find the predicted rate and the tags are recommended to the users. Those results are shown in Table 2

Users	Recommended tags
User 1	Tag 1,Tag 2, Tag3
User 2	Tag 1,Tag 2
User 3	Tag 2,Tag 3
User 4	Tag 1,Tag 4

4. Experimental Analysis and Results

Dataset

The study gathered user-tag relationships and ratings data from a Social Tagging dataset (Delicious). Delicious is a well-known social e-learning platform enabling users to discover new educational materials and organize their bookmarks using keywords. Researchers and developers in the field of collaborative filtering frequently employ it. Matrix notation was used to represent the dataset, with rows standing in for tags, columns for users, and matrix entries for tag weights representing user ratings of resources.

Performance Metrics

The proposed CRS-CF recommendation system was implemented and compared to the classical CF algorithm. This paper contrasts the enhanced slope one method with the suggested hybrid item-based collaborative filtering algorithm. To evaluate the effectiveness of the CRS-CF system, it has been applied to real-world data and compared to the traditional CF algorithm. For this purpose, this study examines the outcomes of various performance indicators, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Precision, Recall, True Positive Rate vs. False Positive Rate, Accuracy, coverage, and user retention rate.

Comparing CRS-CF with Classical CF

The proposed research evaluates the effectiveness of Covering Rough Set-Based Collaborative Filtering (CRS-CF) compared to Classical Collaborative Filtering (Classical

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

CF), two collaborative filtering approaches, by comparing their abilities to generate accurate recommendations based on the following key metrics.

a) Mean Absolute Error

The MAE was a crucial metric for gauging the recommendations' quality. Increased forecast accuracy corresponds to a more petite MAE. The formula for determining this value is as follows:

$$MAE = \frac{1}{n}\sum_{i-1}^{n}|p_i - q_i|$$

Where p_i is the predicted value and q_i is the true value. The empirical results showed that CRS-CF performed exceptionally well, with a remarkable MAE as demonstrated in Fig.1. Compared to the MAE of the traditional CF method, CRS-CF's ability to provide more accurate and trustworthy guidance stands out as the evident benefit.

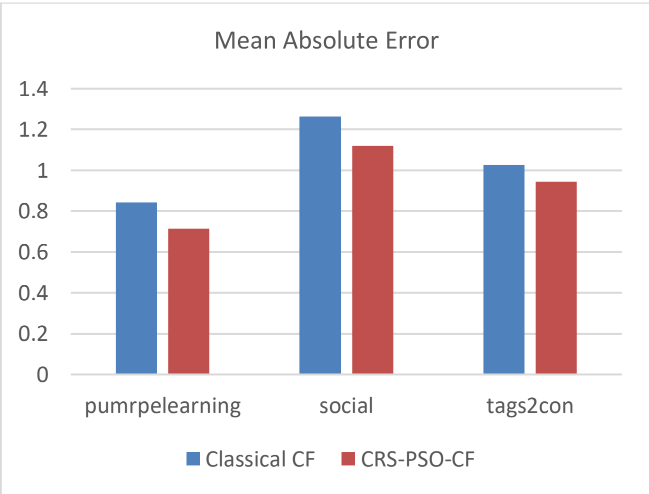

Fig 1. Mean Absolute Error

b) Root Mean Square Error

RMSE is the difference between forecast and corresponding observed values are each squared and then averaged over the sample. RMSE can be calculated as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Where n is the total number of users. y_i is the predicted rate and $\bar{y}_i$ is the actual rate. The comparison in Fig. 2 shows that CRS-CF is superior to classical CF since it has a much smaller RMSE value. This result demonstrates how highly predictive CRS-CF is, which bodes well for its ability to improve recommendation quality significantly. As a result, it aids in developing a superior recommendation system that is also easier to use.

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

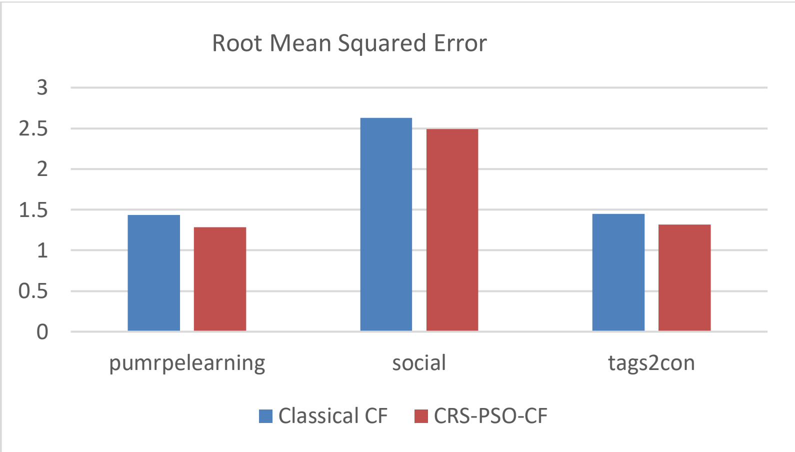

Fig 2. Root Mean Square Error

To further illustrate the performance of our proposed model, we compared the results with the classic CF approach. The following table shows the comparison between CRS-CF and the improved Classical CF algorithm based on the MAE and RMSE metrics. Table 4 shows that the proposed CRS-CF has minimum Mean Absolute Error and Root Mean Square Error when compared with the Classical CF algorithm.

Table 4: Performance Metrics

	Pumrpelearning		Social		Tags2con	
Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE
Classical CF	0.8425	1.438	1.264	2.627	1.027	1.449
CRS-CF	0.7159	1.287	1.119	2.492	0.946	1.319

c) Precision

It is defined as the ratio of the number of recommended objects collected by users appearing in the test set to the total number of recommended objects. This measure is used to evaluate the validity of a given recommendation list. The precision can be formulated as, in which represents the number of recommended products collected by users appearing in test set, and is the total number of recommended products.

$$Precision = \frac{|user\ tags\ \cap recommended\ tags\ |}{|recommended\ tags|}$$

Compared to classical CF, which had a poor precision rate (as shown in Fig. 3), the CRS-CF system displayed outstanding precision, indicating that its proposals were effective.

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

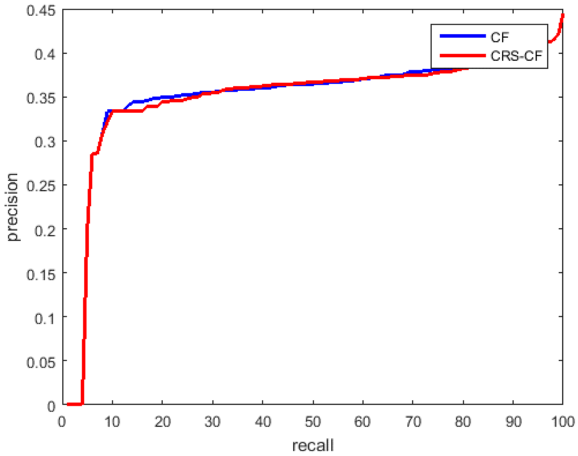

Fig 3. Precision and Recall

d) Recall

It is defined as the ratio of the number of recommended objects collected by users appearing in the test set to the total number of the objects actually collected by these users. The larger recall corresponds to the better performance. The Recall can be formulated as, in which represents the number of recommended products collected by users appearing in test set, and is the total number of these users' actual buying.

$$recall = \frac{|user\ tags\ \cap recommended\ tags\ |}{|favourite\ tags|}$$

CRS-CF (shown in Fig. 3) performed exceptionally well in the rigorous evaluation, as indicated by its high recall score (that means its ability to identify a large proportion of relevant objects accurately). The difference in memory performance between the two recommendation systems is shown by the lower recall rate achieved by conventional CF.

e) True Positive Rate Vs False Positive Rate

TPR is a vital metric for measuring the rate of correctly identified relevant things, while FPR measures the rate of incorrectly identified non-related items. The relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) can be represented as follows:

$$TPR\ =\ (True\ Positives)\ /\ (True\ Positives\ +\ False\ Negatives)$$

$$FPR\ =\ (False\ Positives)\ /\ (False\ Positives\ +\ True\ Negatives)$$

As demonstrated in Fig. 4, the TPR vs. FPR curve for CRS-CF exposed exceptional results, indicating its efficacy in differentiating important from non-relevant elements. On the contrary, the curve for conventional CF firmly confirms the particular performance features that set CRS-CF apart in its capacity to discern between relevant and non-related objects.

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

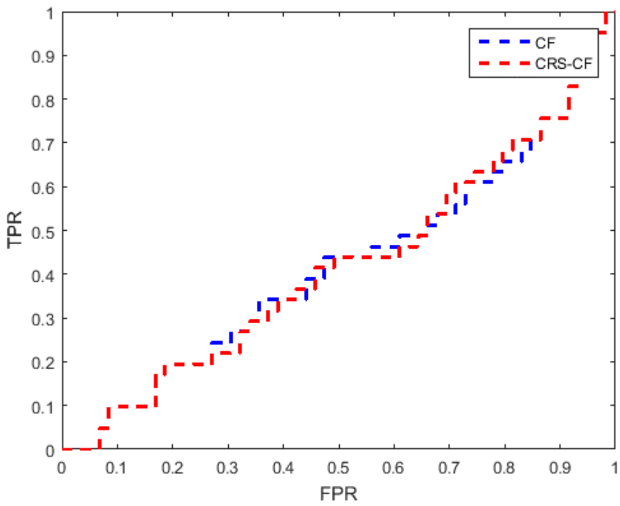

Fig 4. True Positive Rate vs False Positive Rate

f) Accuracy

Accuracy is a crucial parameter in recommendation systems since it measures how often the system makes accurate suggestions based on user preferences and behaviour. Following is the formula for determining a precision rate:

$$\text{Accuracy} = \frac{(\text{True Positives} + \text{True Negatives})}{(\text{True Positives} + \text{True Negatives} + \text{False Positives} + \text{False Negatives})}$$

Fig. 5 shows that CRS-CF achieved an excellent accuracy score, indicating that many of its recommendations were spot on. Classical CF, on the other hand, has a lower rate of success. This distinct distinction highlights the excellent recommendation-making skills of CRS-CF.

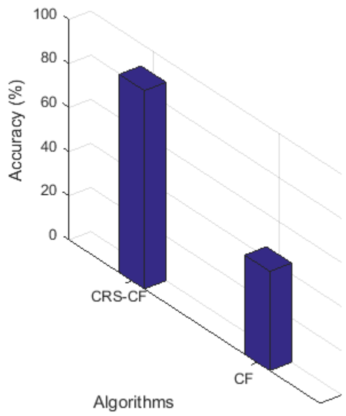

Fig 5. Accuracy of CF Algorithms

g) Coverage Rate

As an essential gauge of the comprehensiveness of a system's suggestions, the coverage rate was the focus of our investigation. Many options that appeal to a wide

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

range of consumers' likes and preferences indicate a high coverage rate. The formula for determining the percentage of people covered is as follows:

$$Coverage\ Rate = (Number\ of\ Recommended\ Items\ /\ Total\ Number\ of\ Items) * 100$$

Fig. 6 shows how the CRS-CF model improves recommendation systems by providing better coverage than Classical CF. The model's wide selection of offerings appeals to content diversification and user satisfaction frameworks since they may meet the needs of a wide range of users. In contrast, the coverage rate for Classical CF is significantly lower. The CRS-CF model's extensive coverage improves the user experience and makes the recommendation system more exciting and tailored to the individual by providing a wide variety of options. Based on the results of this comparison, CRS-CF is the superior option for systems that want to maximize content diversity and user engagement.

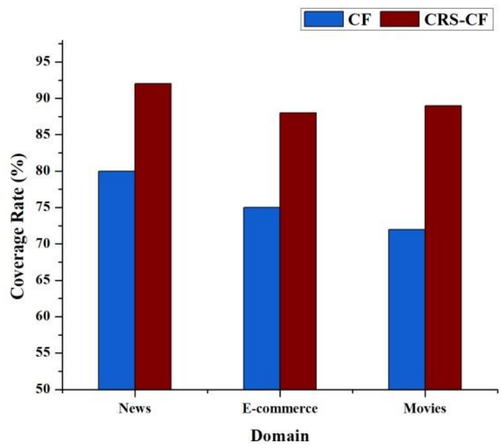

Fig. 6 Coverage Rate (%) of CRS-CF and CF algorithm

h) User Retention Rate

A recommendation system's long-term success and user involvement can be primarily gauged by its retention rate. To determine what percentage of users return, apply this formula:

$$User\ Retention\ Rate = [(E - N)\ /\ S] * 100$$

The variable E represents the number of users after a specific period; new users are represented by N; and the number of users at the beginning is represented by S. According to the results, recommendation systems can sustain users' attention and participation over time. According to Fig. 7, the CRS-CF is the most efficient recommendation system since it keeps a more significant proportion of its customers. These methods, like the Covering Rough Set model, improve customization and precision to suit each user's unique tastes better. Users are more likely to keep using it because of this confidence in and interest in the system. However, the retention rate was slightly lower in classical CF because of the way that it used collaborative filtering in the past.

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

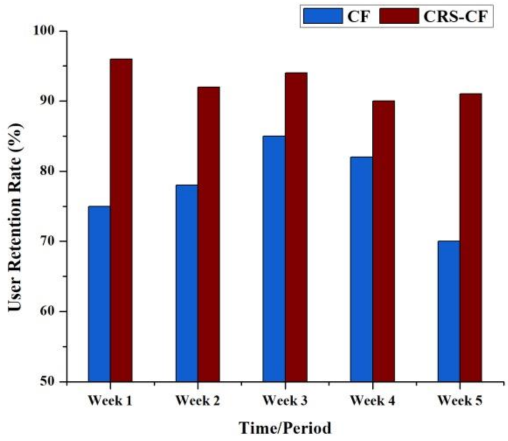

Fig. 7 User Retention Rate (%) of CRS-CF and CF algorithm

The results show that CRS-CF is superior to Classical CF in terms of performance. CRS-CF provides trustworthy recommendations, which increases user happiness and adoption. Accuracy and recall measures improve the quality of requests made to each user. It sorts valuable information from irrelevant data, stimulating the consumer's curiosity. Its high coverage rate can accommodate a wide range of individual tastes, and its low user attrition rate guarantees consistent engagement. Based on these findings, CRS-CF can significantly improve existing recommendation systems by making better-tailored recommendations to each user.

5. Conclusion and Future Work

Recommendation systems play a pivotal role in guiding users to discover content and items of interest by offering tag recommendations, calculated by assessing user similarities. Collaborative Filtering (CF) remains the bedrock of recommendation technology, serving as a well-explored and widely used approach in recommendation systems. In this study, we introduced a novel model, the CRS-CF, which represents a pioneering effort from the perspective of personalization. To the best of our knowledge, our work marks an innovative direction in the realm of Social Tagging Dataset recommendation systems. The CRS-CF model takes an innovative approach by incorporating user-provided weightings for tag items, facilitating a richer and more tailored user experience. Our experimental results demonstrate that the CRS-CF model outperforms Classical CF, underscoring its efficacy in providing more accurate and personalized recommendations. As we look toward the future, our research endeavors will continue to evolve. Specifically, we plan to explore the integration of Uncertain Neighbors to further enhance the covering-based rough set methodology. By optimizing the generation of candidates for recommendation systems, we aim to take our CRS-CF model to the next level, offering even more precise and tailored recommendations. This path of exploration aligns with the ever-evolving landscape of recommendation systems,

Journal of Artificial Intelligence and Data Science Techniques
ISSN: 3029-2794
Volume 01 Issue 01 (February) 2024

where the pursuit of enhanced personalization and recommendation quality remains a driving force. We are committed to contributing to the advancement of recommendation technology, ensuring users benefit from the most relevant and engaging content recommendations in an era of expanding digital resources.

5. References

[1]. Areeb, Q. M., Nadeem, M., Sohail, S. S., Imam, R., Doctor, F., Himeur, Y., ... & Amira, A. (2023). Filter bubbles in recommender systems: Fact or fallacy—A systematic review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, e1512.
[2]. Zhang, Z., Patra, B. G., Yaseen, A., Zhu, J., Sabharwal, R., Roberts, K., ... & Wu, H. (2023). Scholarly recommendation systems: a literature survey. Knowledge and Information Systems, 1-46.
[3]. Pawlak, Z. (1982). Rough sets. International journal of computer & information sciences, 11, 341-356.
[4]. Zhu, W. (2009). Relationship among basic concepts in covering-based rough sets. Information Sciences, 179(14), 2478–2486. doi:10.1016/j.ins.2009.02.013
[5]. Zhu, W. (2009). Relationship between generalized rough sets based on binary relation and covering. Information Sciences, 179(3), 210–225. doi:10.1016/j.ins.2008.09.015
[6]. Zhu, W. (2007). Topological Approached to Covering Rough Sets. Information Sciences, 177, 1499–1508.
[7]. Zhu, W., & Wang, F.-Y. (2012). The fourth type of covering-based rough sets. Information Sciences, 201, 80–92. doi:10.1016/j.ins.2012.01.026
[8]. Zhang, Z., Kudo, Y., & Murai, T. (2015). Applying covering-based rough set theory to user-based collaborative filtering to enhance the quality of recommendations. In Lecture Notes in Computer Science (pp. 279–289). Cham: Springer International Publishing.
[9]. Campagner, A., Ciucci, D., & Denœux, T. (2022). Belief functions and rough sets: Survey and new insights. International Journal of Approximate Reasoning, 143, 192-215.
[10]. Thuan, N. D. (2009). Covering rough sets from a topological pointof view. International Journal of Computer Theory and Engineering, 606–609. doi:10.7763/ijcte.2009.v1.98
[11]. Yang, T., & Li, Q. (2010). Reduction about approximation spaces of covering generalized rough sets. International Journal of Approximate Reasoning: Official Publication of the North American Fuzzy Information Processing Society, 51(3), 335–345. doi:10.1016/j.ijar.2009.11.001
[12]. Sachan, A., & Richariya, V. (2013). A Survey on Recommender Systems based on Collaborative Filtering Technique. International Journal of Innovations in Engineering and Technology, 2(2).
[13]. Babu, R. V. (2010). A New Approach for Cluster Based Collaborative Filters. International Journal of Engineering Science and Technology, 2(11), 6585–6592.
[14]. Zakowski, W. (1983). Approximations in the Space (u, π). Demonstration, 16, 761–769.
[15]. Tsang, E. C. C., Chen, D., Lee, J. W. T., & Yeung, D. S. (2005). On the upper approximations of covering generalized rough sets. Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826). IEEE.
[16]. Wang, J., Dai, D., & Zhou, Z. (2004). Fuzzy Covering Generalized Rough Sets. Journal of Zhoukou Teachers College, 21, 20–22.