Sushma Malik et al. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 10, Issue 3 September 2023, pp. 27-35

A New Hybrid Approach to Resolve the Gray Sheep Users Problem In Recommender Systems

Sushma Malik, Dr. Mamta Bansal

School of Engg. & Tech. Dept. of CSE, Shobhit Institute of Engineering & Technology, Meerut (250110), India

Abstract: **The success of e-commerce has strengthened traditional trading patterns. Instead of physically visiting stores, many now find it more convenient to do their shopping online. Users are having difficulty getting information for webs due to the expansion of the Web and the uncontrolled enormous amount of information. Recommender systems are successfully implemented in the digital platform to eliminate the information overload problem and also assist the user in the decision-making process. To improve the system performance and prediction accuracy it is a vital task to identify the Gray-sheep users from the white users of the system. In this paper, to overcome this problem, a new approach called PowKmeans++ has been introduced to identify the gray-sheep users by using the clustering approach of data mining and provides a recommendation to the gray- sheep users. The convenience and effectiveness of the proposed approach are demonstrated by our experimental findings based on KaggleBooks rating data. . We compare the PowKmeans++ result with KMeans++ and observed that PowKmeans++ performs better than KMeans++ and we are using the MAE and F-score metrics to show the result.**

Keywords: **Recommender system (RS), Collaborative Filtering (CF), Gray-sheep user, White user, KMeans++, PowKmeans++ .**

I. Introduction

Nowadays, the communication technologies like the internet and electronic services have increased the generated binary data and transfer of information. However, as the number of internet users grows, so does the amount of digital data stored on these networks, culminating in a problem of information overload[1][2]. With the current information overload problem, filtering and efficiently providing relevant information to users based on their requirements and interests has become extremely tough. This information overload problem highlights the necessity for an information extraction system that can extract unseen data and forecast whether a user will appreciate a particular preference. Recommender systems(RS) are an example of this type of system[3][4]. RS can make product recommendations depending on the user's interests and preferences. Such preferences could be retrieved either explicitly or implicitly[5]. Therefore, the use of content-based RS becomes a need and single solution to solve the information overload issue on the digital platform. RS is widely used in many domains like E-Commerce[6], digital libraries[7], health[8], tourism[9], marketing[10] and many more. To fulfill the needs of the user, many electronic portals are using the RS. Table 1 shows some RS that are used by various sites.

RS depends on several factors like ratings to items or products given by the user on their satisfaction level, users like and dislike, age, gender, location, locality, community, and many more (Jindal et al., 2018). E-Commerce portals get many benefits from using the RS like changing browsers into buyers; rising sales and also increasing the loyalty of the user towards the site. Increase the sales of the sites by providing the recommendation of new products to the customer on their taste and requirement. Based on the user's shopping history, RS will automatically recommend products to them.

Table 1: List of Some Recommender Systems

Site	Recommended Products or Services
Amazon.com	Products
Facebook.com	Friends
MovieLens.org	Movies
Spotify	Songs
YouTube.com	Videos
Netflix.com	Movies
Twitter.com	Friends
Foursquare.com	Locations
Yelp.com	Restaurants
Google.com/maps	Routes
ScienceDirect.com	Articles
Kaggle.com	Books

For example, a user might show interest in a product by reading its description on the website or adding it to their shopping cart. The consumer will then be given recommendations for relevant products. Sometimes products are also recommended based on top overall sellers on a site[11][12].

The RS is designed using the most generally utilized recommendation technique, CF. This technique is based on the collection and analyzing an enormous amount of information based on users' behaviors, their activity on sites, and predicting the products by analyzing similar user's preferences. It is predicated on the assumption that consumers who previously agreed will agree again in the future and will wish to acquire similar products to those they previously bought or loved[13]. With the benefit of the CF technique in the RS, it also introduced some problems like cold-start, data sparsity, and gray-sheep user problems[14][15]. When a user's purchase history is not available in the site database, RS faces a cold start difficulty and fails to recommend products. The problem of data sparsity arises due to the vast size of the user-item matrix, which is exceedingly sparse. Gray-sheep user is also a problem that occurs in the CF technique[16][17]. For users who are not having similar behavior to other users and are not able to relate with a particular group is tough to give an accurate recommendations are called grey-sheep users. This paper has focused on the problem of these types of users for the recommendation. A user with a high correlation with other users is called the white user otherwise a Gray-user with no correlation with other users. At that time, RS is not predicting users' preferences with this CF technique problem. The occurrence of gray-sheep users in the RS has mainly two drawbacks[14]:

1. These users cannot obtain an accurate recommendation.
2. Also, affect the white user's recommendation process.

Separating gray-sheep users from white-sheep users is a good technique to lessen gray-sheep users' influence on RS performance. As a result, we introduced PowKmeans++, and unique method for identifying grey-sheep users from active users. The hybrid RS is then further designed to provide a recommendation to these identified grey-sheep users.

II. Literature review

Table 2: Literature Review of Gray-sheep users in RS

S.No	Reference	Review
1	[18]	Grey sheep dilemma is a challenge for recommendation systems that needs to be overcome with a hybrid approach.
2	[19] [20]	The "grey sheep" issue is unique to pure collaborative filtering methods, since the feedback provided by one person does not correspond to any user neighbourhood. In this case, the system is unable to correctly predict items that are pertinent to that user. Pure content-based techniques, where predictions are based on the user's profile and item qualities, can overcome this challenge.
3	[21]	Recommended a fresh CF strategy to address the Gray sheep problem. With the help of users with various interests and preferences, this strategy seeks to grow the active user's number of neighbours.
4	[12]	The Gray-sheep one-class recommendation (GSOR) framework was presented for identifying gray-sheep users and making recommendations to them using hybrid RS.
5	[1]	Introduced the new algorithm named MKMeans++ by adding power weight and power item feature with KMeans++ algorithm to identify the gray-sheep users.
6	[22]	Using histogram intersection to provide greater user-user similarities to get better recognition of gray-sheep users.
7	[11]	Introduced PowKMeans algorithm to detect the gray-sheep users and also provides the recommendation to these identified users.
8	[23]	Proposed a new CF technique to handle the grey sheep problem. The main motives to improve the accuracy of prediction are by turning the users whose preferences are not matched with the target user into new neighbors.
9	[24]	Design the MF-based RS that improves the recommendation accuracy to the grey users.
10	[25][26]	According to the author, GSU is blamable to increase the fault rate in CF-based RS. To detect these kinds of users, the clustering method is used and the recommendation of the products to GSU on the support vector machine regression.
11	[27]	Introduced the novel distribution-based technique to identify the GU that borrows from outlier detection and information retrieval.
12	[3]	CF technique is not giving an accurate recommendation to the GSU. The author introduced a new K average cluster solution that identifies these users and has created a recommendation for the user deployment of these types of users.
13	[28][29]	According to the writers, there are three sorts of users on E-Commerce sites: white, black, and gray users who provide reviews, feedback, and are more engaged with the system which helps the RS have adequate data for those users which help to create an association with other users with same taste are called white users. On the other hand, users who access the sites but do not share their comments and feedback which causes problems to find the relationship of those users with other users are called black

		users. Grey sheep users have unusual preferences which create less correlation with other users and affect the functioning of RS.
14	[30]	Repurposed outlier detection strategies based on the distribution of user similarities to other users to provide new ways to aid in the recognition of the GSU.
15	[31]	For ensuring the efficiency of RS which is designed on the CF technique, detection of GSU is the imperative task in the process of recommendation. The author proposed and empirically showed that a personality-based user model may be utilized to identify GSU, as well as that deleting GS users, can improve the overall efficiency of CF-based RS.

Gray-sheep Users

As E-Commerce sites have grown in popularity, a growing number of users with varying tastes have turned to the recommendation system to meet their demands and expectations. RS enables users to purchase their selected product on-site in a short time, saving time and without the need to physically visit their preferred store. The most common creative strategy is CF, which is used to create recommendations for consumers similar to other users on the system. The key expectation of the CF technique is that a user will choose the same product that is purchased or liked by other users who have similar tastes as the target user.

Table 3: Us2 Preferred Products List

Users	**Product Preferred**
U1	P2, P3, P6
U2	P1, P3, P6, P5
U3	P4

For example in RS, U1, U2, and U3 are the three users where P1, P2, P3, P4, P5, and P6 are the available products on the site. As presented in table 3, user U1 likes the P2, P3, and P6, P1, P3, and P6 are liked by the user U2. On the other side, user U3 likes the product P4 to purchase. There is a strong probability that U1 will prefer the P1 because users U1 and U2 have the same taste as they prefer the same products (P3 and P6). User 1 and User 2 cannot get the preference of P4 product as there is no similarity of taste between users U1, U2 with user U3. In the same way user, U3 cannot get the recommendation of products P1, P2, P3, P5, and P6 because the profile of user U3 does not match with other users like U1 and U2. Such users in the RS whose profile is not matched with other users and RS helpless to provide the recommendation of products to such users and these users are called the gray-sheep users. Large numbers of gray-sheep users are present in the CF-based recommender system and affect the recommendation process. The presence of the gray-sheep user in RS has mainly two effects these kinds of users are not getting the accurate recommendation as per their needs and also affect the recommendation process of other users (white users). Separating gray-sheep users from white users is an effective method for RS to improve its functionality. Data mining's clustering strategy is the most effective tool for separating these users. Users with strong correlations are grouped and placed in one cluster, whereas users with low correlations are clustered together and placed in another cluster.

III. Proposed approach

The architecture of the proposed recommendation system for the identified gray sheep users is shown in figure 1. It consists of the three main steps:

1. User item Rating Matrix to extract features and introduce the CentroidSelect algorithm.
2. Detection of Gray-sheep user
3. Recommendations to identify Gray-sheep users.

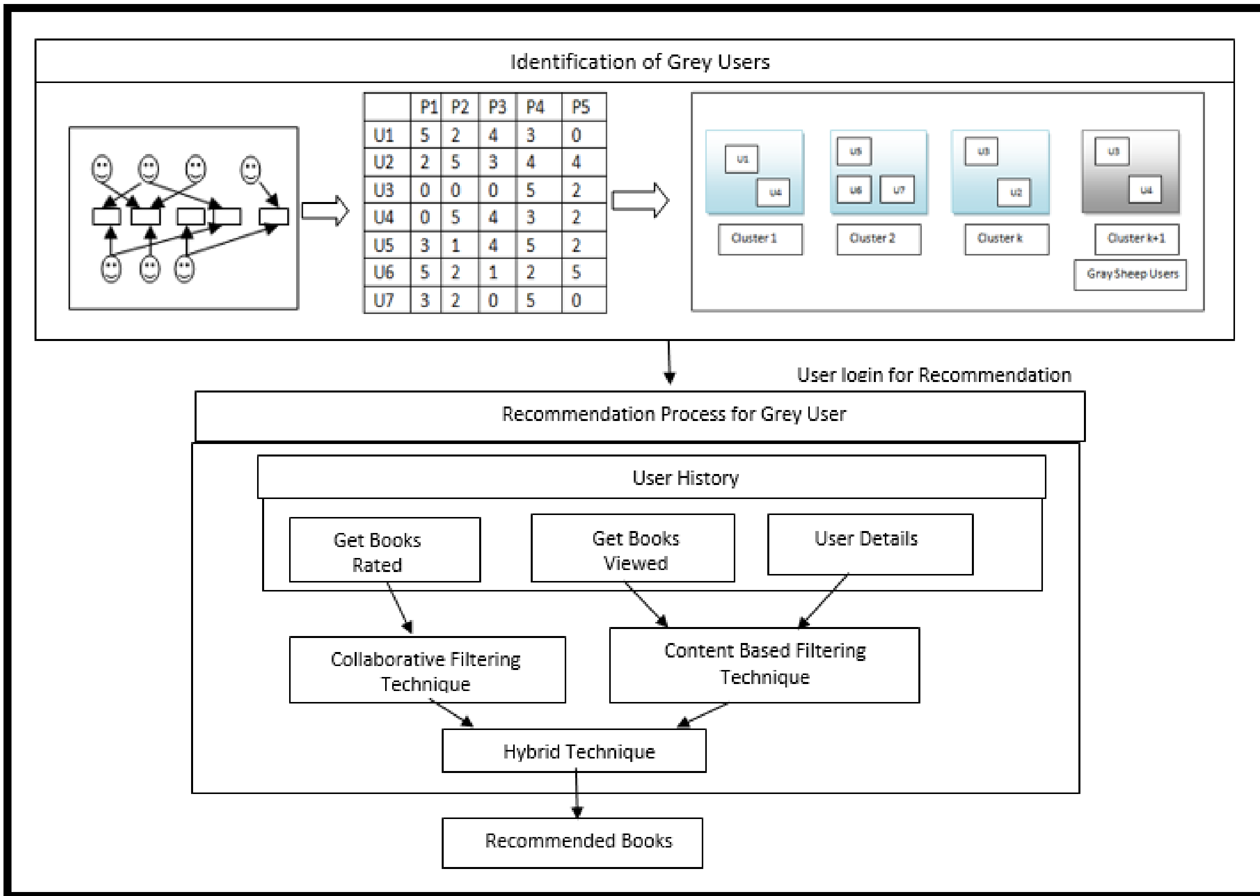

Figure 1: RS complete architecture

Step 1: User item Rating Matrix to extract features and introduce the CentroidSelect algorithm.

A user-item rating matrix is created in this stage, with rows representing the number of users (U) and columns representing the Items (I). The value of (U, I) denotes the ratings of users, which range from 1-5, while the non-rated item is assigned a value of zero (0). The rating matrix of the CF approach contains two categories of consumers: white users and GSU. As a result, the quality of the recommendation to the other group is influenced by data from each group. As a result, the main motivation is to separate gray sheep users. One of the available methods for distinguishing gray sheep users from white sheep users is the KMean clustering algorithm. Users are set initially in the n clusters at first, and afterward, the users with the best separation from the cluster's centroid are put in another group. This partitions white clients into k groups, while gray-sheep users have put in the (k+1)th cluster. The KMean clustering algorithm initially selects k centroids randomly and the KMeans++ clustering algorithm is introduced (Vassilvitskii & Arthur, 2007) which uses the distance features to select the initial cluster centroids. Distance is calculated by eq. (1), where D (x_i) is the minimum distance between the nearest centroid and x_i. P_{matrix} is the user set called preference matrix.

$$P_{matrix}\ (i) = \frac{D\ (x_i)2}{\sum_{xiEu} D\ (x_i)\ 2} \qquad (1)$$

In this paper, a customized version of the KMeans++ clustering method is used to cluster the user-item rating matrix. We extend the KMeans++ method to include users having a significant role in the rating matrix are power users with a new approach, PowKmeans++, is provided. We choose power users as the starting centroids in this case because it may improve convergence and increase the quality of the resulting clusters.

The term power user (Up) can be defined by equation (2) as follows:

$$Up = \frac{|I_u^r|}{|I_u|} \qquad (2)$$

In equation (2), Up represent the power user, I_u represents the maximum number of items that are rated by a user and I^r_u represents the number of data items rated by user u.

The number of normalized ratings has been used in place of normal ratings as per equation (3).

$$P\,(u) = \frac{|I_u|\ (\textit{Nuber of items rated by users})}{|\,I_{up}|\ (\text{Items rated by Power Usres})} \qquad (3)$$

The things with the highest number of users who rated them are referred to as power items, whereas weak items had the lowest number of users who evaluated them. Equation (4) can be used to express the power item as follows:

$$PI = \frac{|u_i|}{|i_p|} \qquad (4)$$

In equation (4), the maximum number of users who rated an individual data item is represented by |Ip|, and the number of users who have rated the item is represented by |ui|. The value obtained from equation (4) is between zero (0) and one (1), displaying how popular each item is in comparison to the most popular goods. I and Pi represent the power item value for item i.

Distance Measure

The distance function is used to calculate the distance between the centroid and the user. In this scenario, the Pearson Correlation likeness between the user and centroid is used. When the similarity is negligible i.e. zero (0), the distance between them is greatest, and vice versa[25]. Equation (5) uses the following simple equation to explain the distance function:

$$\text{dist}\,(u) = \frac{\frac{1}{Sim(u)}}{Max_{Dist}} \qquad (5)$$

The maximum distance between two users is represented by MaxDist (set to 1000). The Pearson correlation can have a negative similarity between two users, which means it can't be utilized to compute distance in equations (5). To circumvent this, we add one (1) to all Pearson correlation similarities before utilizing them in equations (5).

$$\text{Sim}\,(u) = Sim_{(u)} + 1 \qquad (6)$$

Sim (u) in equation (6) shows the Pearson correlation value.

Centroid Selection Approach

Algorithm 1: CentroidSelect, select k users as centroids from the dataset.

Input: u (user), k (Number of clusters), p (u) (Normalized ratings)

Output: {c1, c2,………ck}

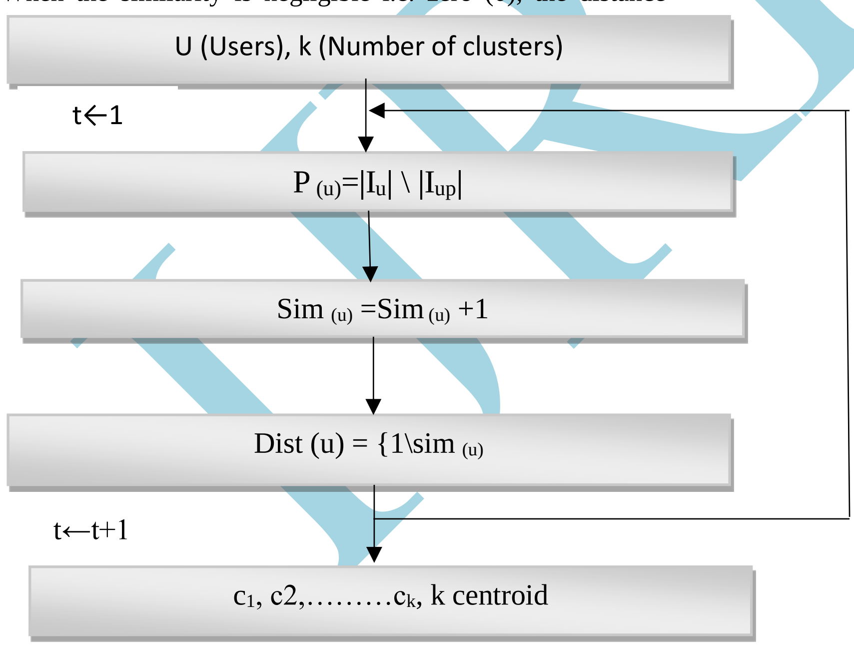

Figure 2: Flow chart of CentroidSelect Algorithm

Step 2: Detecting Gray Sheep Users

By evaluating the user's similarity to the centroid, each user is grouped into clusters. Users with a high degree of similarity are placed in the same cluster, while those with a high degree of dissimilarity are placed in a distinct cluster. Users that have no resemblance to other users are placed in their cluster. The white sheep users are represented by the first k clusters, whereas the grey sheep users are represented by the last (k+1) cluster.

Gray sheep users have very less preferences in common with other users, and their distance from their clusters of centroids is larger than remaining users in the same clusters[32]. The gray sheep users are separated using a distance criterion. As a result, we need to set a distance/similarity criterion to distinguish the gray-sheep users by introducing a threshold value, which ranges from +1 to -1 which is used to calculate the correlation coefficient between two users.

Users with a higher similarity threshold value (+1) are placed in the gray sheep group, while those with a lower similarity threshold value (-1) are placed in the white sheep group. We should choose an appropriate value for the similarity threshold that falls between +1 and -1.

$$\text{Dist (u)} = 1 - Thresholdsimilarity_{(u)} \quad (7)$$

Algorithm 2: GSUSelect, Users' item ratings are grouped into a k+1 cluster, with Gray-sheep users in a distinct cluster.

Input: u (users), k (Number of the cluster), α (A threshold for detecting grey sheep users based on similarities)

Output: ClusterW (White users) and ClusterG (Gray-sheep users)

Obtain k clusters from CentroidSelect Algorithm

↓

Each user's distance from the cluster centre to which they belong was measured.

↓

The user will be placed in the gray sheep user cluster if the user distance exceeds the ideal threshold value; else, it will be placed in the whitesheepuser cluster

↓

K groups consists all white users while separated Gray-sheep users placed in the $(k+1)^{th}$ cluster.

Figure 3: Flow chart of GSUSelect Algorithm

Step 3: **Recommendation to the gray-sheep users.**

In the hybrid technique where CF and Content-based algorithms are combined to give recommendations to GSU, a collaborative technique is employed to propose books and track down the people who evaluated them. Users are screened out again based on demographic characteristics of clients such as age, gender, and so on. To identify the resemblance between the filtered users and the present user, a content-based method is applied. After that, another filter is applied, this time to propose books based on what the user has seen and how other users have rated them. Hybrid recommendation techniques reduce the number of user comparisons while simultaneously providing the user with relevant recommendations from the recommender system.

IV. EXPERIMENTAL RESULTS

Dataset and Experimental Setup

This research makes use of the KaggleBook dataset, which includes 981756 ratings, 10,000 books, and 53424 users. Each user has given ratings to at least 20 books, with ratings ranging from 1 to 5. Books with a rating of 1 are the least desired by the user, while books with a rating of 5 are the most favored. (http://kaggle.com)

Assessment Metrics

Two statistical measures are utilized to assess the efficiency of the proposed strategy. Mean Absolute Error (MAE) and F1-Score are measures in question.

Mean Absolute Error (MAE)

The average magnitude of the errors in a series of forecasts is known as MAE. It is the average of the absolute discrepancies between forecast and actual observation over the test sample. It can be calculated with Eq.1

$$MAE = \frac{\sum_{i=1}^{N} |Pi-Qi|}{N} \qquad (8)$$

P represents the prediction data and Q represents the actual observation. The motive of RS is to diminish the MAE score.

F-Score

F-score is used to measure the accuracy of the model on a dataset. It is the combination of precision and recall of the model. Recall calculates the number of positive predictions made out of all predictions. So the recall is the fraction of relevant recommendations to the total relevant recommendations. Recall can be expressed as in Eq. 2. Precision can be defined as the fraction of relevant instances to the total instances. The calculation of Precision shown in Eq. 3 and Eq.4 represents the formula of the F-score. The motive of any RS is to achieve a high F-score.

$$Recall = \frac{Truepositive\ (TP)}{TruePositive(TP)+TrueNegative(TN)} \qquad (9)$$

$$Precision = \frac{RelevantInstances}{TotalInstances} \qquad (10)$$

$$F-Score = 2 * \frac{Precision*Recall}{Precision+Recall} \qquad (11)$$

V. Results

Results are given by comparing the results given by the proposed (PowKmeans++) method with the existing method (KMean++) (Rashidi et al., 2020) for detecting gray-sheep users in this section. Figure 4 illustrates the F-score of PowKmeans++ versus KMean++, which clearly reveals that PowKmeans++ is more accurate than KMean++. In comparison to KMeans++, the PowKmeans++ approach has a lower error rate (see Figure 5).

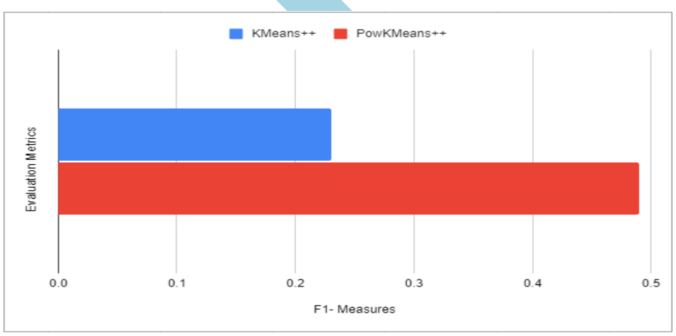

Figure 4: F1-Score in GSU detection

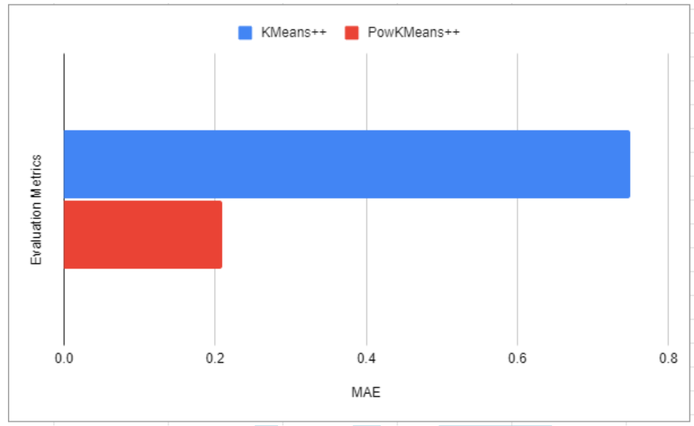

Figure 5: MAE in GSU discovery

VI. Conclusion

RS plays an essential role in the digital platform where lots of information is present for the users and in that situation users face the information overload problem. RS designed with CF techniques has faced a number of problems like data sparsity and gray-sheep users. RS are failed to provide a precise recommendation to GSU. The presence of gray-sheep users in the RS not only affects the recommendation process but also provides a negative recommendation to other users. To overcome this problem, a new approach called PowKmeans++ has been introduced based on CF for detection and provides a recommendation to the gray- sheep users. We compare the PowKmeans++ result with KMeans++ and observed that PowKmeans++ performs better than KMeans++ and we are using the MAE and F-score metrics to show the result.

References

[1] R. Rashidi, K. Khamforoosh, and A. Sheikhahmadi, “An analytic approach to separate users by introducing new combinations of initial centers of clustering,” *Phys. A Stat. Mech. its Appl.*, vol. 551, p. 124185, 2020.

[2] M. S. Ibrahim and C. I. Saidu, “Recommender systems: algorithms, evaluation and limitations,” *J. Adv. Math. Comput. Sci.*, vol. 35, no. 2, pp. 121–137, 2020.

[3] M. A. Ghazanfar and A. Prügel-Bennett, “Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems,” *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3261–3275, 2014.

[4] K. S. A. Z. Ullah, “I Kim K Joukov N Recommender Systems: Issues, Challenges, and Research Opportunities,” *Inf. Sci. Appl.*, vol. 2016, 2016.

[5] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, “Facing the cold start problem in recommender systems,” *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2065–

2073, 2014.

[6] S. P. Singh and S. Solanki, "Recommender system survey: Clustering to nature inspired algorithm," in *Proceedings of 2nd International Conference on Communication, Computing and Networking: ICCCN 2018, NITTTR Chandigarh, India*, 2019, pp. 757–768.

[7] A. Sase, K. Varun, S. Rathod, and D. Patil, "A proposed book recommender system," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 2, pp. 481–483, 2015.

[8] H. M. Habeeb, A. Al-Azawei, and N. Al-A'araji, "Developing a healthcare recommender system using an enhanced symptoms-based collaborative filtering technique," *J. Comput. Theor. Nanosci.*, vol. 16, no. 3, pp. 920–926, 2019.

[9] J. Borràs, A. Moreno, and A. Valls, "Intelligent tourism recommender systems: A survey," *Expert Syst. Appl.*, vol. 41, no. 16, pp. 7370–7389, 2014.

[10] K.-W. Cheung, J. T. Kwok, M. H. Law, and K.-C. Tsui, "Mining customer product ratings for personalized marketing," *Decis. Support Syst.*, vol. 35, no. 2, pp. 231–243, 2003.

[11] H. Jindal, S. Agarwal, and N. Sardana, "PowKMeans: a hybrid approach for gray sheep users detection and their recommendations," *Int. J. Inf. Technol. Web Eng.*, vol. 13, no. 2, pp. 56–69, 2018.

[12] R. Alabdulrahman and H. Viktor, "Catering for unique tastes: Targeting grey-sheep users recommender systems through one-class machine learning," *Expert Syst. Appl.*, vol. 166, p. 114061, 2021.

[13] S. Jaiswal and A. Ajgaonkar, "Online shopping recommendation optimization based on user's previous search history," in *Student, MCA Department, Sardar Patel Institute of Technology*, vol. 3, no. 04, 2016.

[14] G. Bejaoui and R. Ayachi, "A trust-based clustering approach for identifying grey sheep users," in *Digital Economy. Emerging Technologies and Business Innovation: 5th International Conference on Digital Economy, ICDEc 2020, Bucharest, Romania, June 11–13, 2020, Proceedings 5*, 2020, pp. 76–88.

[15] S. Puntheeranurak and H. Tsuji, "A multi-clustering hybrid recommender system," in *7th IEEE International Conference on Computer and Information Technology (CIT 2007)*, 2007, pp. 223–228.

[16] D. Sánchez-Moreno, A. B. G. González, M. D. M. Vicente, V. F. L. Batista, and M. N. M. García, "A collaborative filtering method for music recommendation using playing coefficients for artists and users," *Expert Syst. Appl.*, vol. 66, pp. 234–244, 2016.

[17] A. Srivastava, "Gray sheep, influential users, user modeling and recommender system adoption by startups," in *Proceedings of the 10th ACM conference on recommender systems*, 2016, pp. 443–446.

[18] P. P. Rokade, P. V. R. D. P. Rao, and A. K. Devarakonda, "Forecasting movie rating using k-nearest neighbor based collaborative filtering," *Int. J. Electr. Comput. Eng.*, vol. 12, no. 6, pp. 6506–6512, 2022, doi: 10.11591/ijece.v12i6.pp6506-6512.

[19] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *J. Big Data*, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5.

[20] C. T. Gadkari, "RECOMMENDATION SYSTEM AND IT ' S TYPES," no. 02, pp. 534–549, 2023.

[21] A. El Fazziki, Y. E. M. El Alami, J. Elhassouni, O. El Aissaoui, and M. Benbrahim, "Employing opposite ratings users in a new approach to collaborative filtering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 1, pp. 450–459, 2022, doi: 10.11591/ijeecs.v25.i1.pp450-459.

[22] Y. Zheng, M. Agnani, and M. Singh, "Identification of grey sheep users by histogram intersection in recommender systems," in *Advanced Data Mining and Applications: 13th International Conference, ADMA 2017, Singapore, November 5–6, 2017, Proceedings 13*, 2017, pp. 148–161.

[23] A. El Fazziki, Y. E. M. El Alami, J. Elhassouni, O. El Aissaoui, and M. Benbrahim, "Employing opposite ratings users in a new approach to collaborative filtering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 1, pp. 450–459, 2022.

[24] B. Gras, A. Brun, and A. Boyer, "Can Matrix Factorization Improve the Accuracy of Recommendations Provided to Grey Sheep Users?," in *13th international conference on Web Information Systems and Technologies (WEBIST)*, 2017, pp. 88–96.

[25] M. Ghazanfar and A. Prugel-Bennett, "Fulfilling the needs of gray-sheep users in recommender systems, a clustering solution," 2011.

[26] S. Ghorbani and A. H. Novin, "An introduction on separating gray-sheep users in personalized recommender systems using clustering solution," *Int. J. Comput. Sci. Softw. Eng.*, vol. 5, no. 2, p. 14, 2016.

[27] B. Gras, A. Brun, and A. Boyer, "Identifying grey sheep users in collaborative filtering: a distribution-based technique," in *Proceedings of the 2016 conference on user modeling adaptation and personalization*, 2016, pp. 17–26.

[28] P. Kumar and R. S. Thakur, "Recommendation system techniques and related issues: a survey," *Int. J. Inf. Technol.*, vol. 10, pp. 495–501, 2018.

[29] A. Tennakoon, N. Gamlath, G. Kirindage, J. Ranatunga, P. Haddela, and D. Kaveendri, "Hybrid recommender for condensed sinhala news with grey sheep user identification," in *2020 2nd International Conference on Advancements in Computing (ICAC)*, 2020, vol. 1, pp. 228–233.

[30] Y. Zheng, M. Agnani, and M. Singh, "Identifying grey sheep users by the distribution of user similarities in

collaborative filtering," in *Proceedings of the 6th annual conference on research in information technology*, 2017, pp. 1–6.

[31] A. Srivastava, P. K. Bala, and B. Kumar, "New perspectives on gray sheep behavior in E-commerce recommendations," *J. Retail. Consum. Serv.*, vol. 53, p. 101764, 2020.

[32] J. P. Lucas, N. Luz, M. N. Moreno, R. Anacleto, A. A. Figueiredo, and C. Martins, "A hybrid recommendation approach for a tourism system," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3532–3550, 2013.