2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS) | 979-8-3503-9874-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/SCEECS57921.2023.10063125

Hybrid and Classical Models of Recommendation Systems- A Review

Mohd Mustafeez ul Haque
Department of CS & IT
Maulana Azad National Urdu University
Hyderabad, India
mohdmustafeezulhaque@gmail.com

Bonthu Kotaiah
Department of CS & IT
Maulana Azad National Urdu University
Hyderabad, India
kotaiah.bonthu@manuu.edu.in

***Abstract*—Recommendation Systems (RS) have evolved as a response to information overload caused by the growth in online content, increasing users' search time and information retrieval rate. By providing consumers with personalized suggestions based on their past preferences, likes, and dislikes, RS is used to find relevant content. Products, videos, photos, articles, news, and books are a few of the areas where these technologies are applied. Both attribute data, such as textual profiles and significant keywords, and user-item interactions such as ratings or purchasing activity are used by RS to analyze. This article provides an outline of the needs and difficulties of RS. To achieve greater recommendation capabilities in a variety of situations, this study attempts to explain the limits of the current generation of recommendation approaches and potential expansions with recommender systems. The results of the proposed methods have good improvement when compared with existing techniques.**

***Keywords*—Collaborative Filtering, Clustering, Content-Based Filtering, Hybrid Filtering, Ontology, Recommender systems, Sequential pattern mining**

I. Introduction

In order to filter the huge amount of data circulating on the internet, Recommendation Systems (RSs) become necessary and can be exploited by ordinary users. Javed et al.[1] Created a system that can provide suggestions depending on the preferences and requirements of the users. The Recommender System's Taxonomy is depicted in Fig. 1.

A. Content-Based Filtering (CBF)

Without considering any information about other users, CBF forecasts the user's preferences based solely on his information (gender, age, social media interactions, etc.). It is illustrated in Fig 2. Since CBF (Papadakis et al.[2], Sharma et al.[3]) employs a number of procedures to provide the necessary information to the individual who needs it, it may be characterized as an information filtering task. Instead of scanning an incoming stream for particular data, filtering is sometimes thought of as the process of eliminating undesired input, or what is known as noise.

B. Collaborative Filtering (CF)

CF filters the flow of data that can be recommended, by RS, to a target user according to taste and preferences shown in Fig 2. The target user's profile is built based on similarity with other users. According to Sharma et al.[3], Sharma et al.[4], and Javed et al.[1], the approach is particularly sensitive to the similarity measure used to gauge the degree of dependence between two users (or two products). A collection of comparable neighbors is established for each user of a CF, and the choice of whether or not to propose a product to a user is based on the opinions of those neighbors. The memory-based (neighborhood- or heuristic-based) and model-based sub-approaches of CF-based RS[5] may be distinguished.

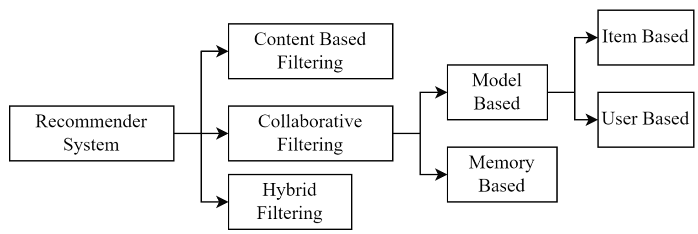

Fig. 1. Classifiction of Recommender System

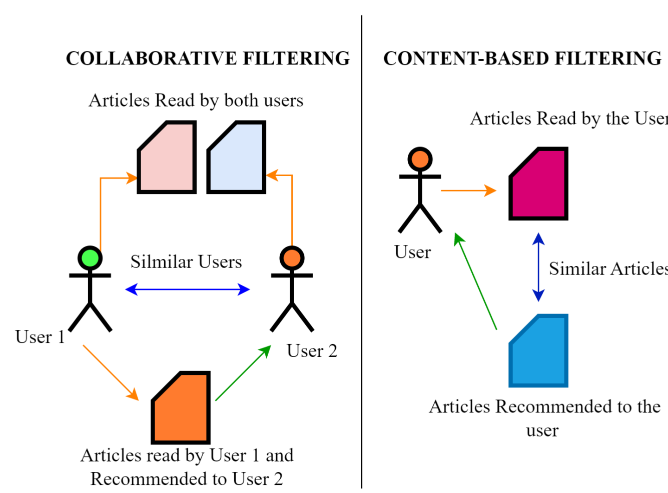

Fig. 2. Collaborative Filtering & Content-Based Filtering

1) Model-based Method: Unlike other methods that use ratings directly in predictions, model-based approaches learn how to predict using these ratings. Model-based systems model interactions between users and items based on underlying factors representing their hidden characteristics, such as priority categories for users and categories for items. Papadakis et al.[2] states that the trained model is used to forecast user ratings for new goods. The model is trained using the data that is currently available (train dataset). Singular Value Decomposition, Latent Semantic Analysis, and Bayesian Clustering are just a few model-based methods.

2) Memory-based Method: For predicting new item classifications in memory-based CF, user-item ratings are stored in the memory and when compared with the model-based approach, this memory-based method is found to achieve better performance in terms of accuracy to handle massive data sets. Won et al.[6] And Sharma et al.[3] Has done the categorization into two classes, user-based and item-based recommendation.

3) User-Based CF: Users in the community who have already evaluated articles will be used to filter the flow of new items using user-based CF. Figure 3 depicts User-Based CF. An item will automatically be recommended to users who have previously expressed similar ideas if it has been deemed interesting by one user. In order to store the user similarity scores[4], [6], the system must create a User's, Users matrix. As a result, the rating a user gives a product is based on the neighbors who are also users of that product.

4) Item-Based CF: The similarity scores between items are kept in a matrix called Items, Items in the item-based CF method. A collection of products that have already been highly evaluated by the active user [4], [6] is presented in Fig 4. The anticipated rating is determined by the degree of similarity between the item and its neighbor; the higher the degree of similarity, the more similar the projected rating is to the neighboring rating.

***Challenges in Collaborative Filtering*:**

Cold Start Problem (CSP): When the catalog items either have lesser interactions or no interactions, the cold-start problem will occur which makes the system not refer any new items that haven't received any user rating. The three cases of cold-start problems are new items, new communities, and new users. [5], [7].

Data Sparsity problem: When there are not enough data available to predict similar users, a sparse problem will occur and the main reason behind this problem is that the users were rating only a small subset of items. This will end up making the ratings sparse and this will affect CF methods as they are depending upon the rating matric always while they are making appropriate recommendations.

Scalability: Due to the vast number of people and items on the web, as mentioned by Ramzan et al.[8] and Yang et al.[9], it takes a lot of processing resources to determine how similar users are in order to provide suggestions.

C. Hybrid filtering

It is possible to create a hybrid strategy that addresses the drawbacks of both collaborative filtering and content-based strategies. To improve performance, Sharma et al.[4] merged one or more recommendation approaches, as seen in Fig 5. By looking at their profiles, the first stage identifies people who are comparable to the current user. The cosine equation is used to determine similar users based on the user's ratings. Once this computation is complete, find out which users are neighbors or friends of the top n. In the second phase, the candidate's item for each neighbor is selected. After determining each item's prediction value, the third step suggests the items to the target user.

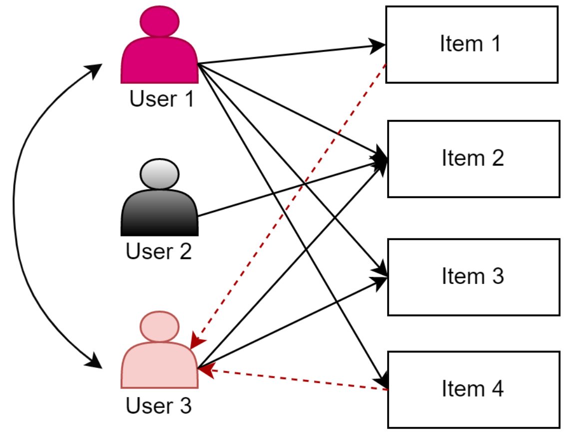

Fig. 3. User Based CF

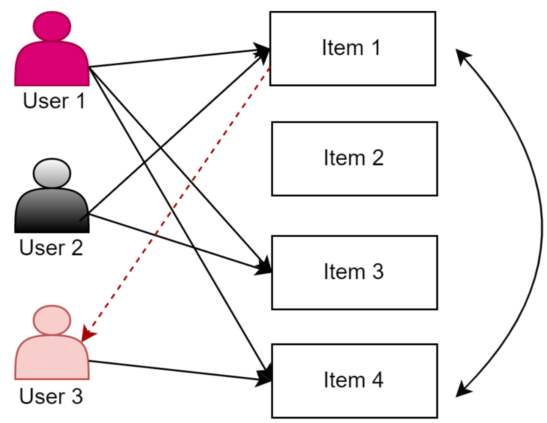

Fig. 4. Item Based CF

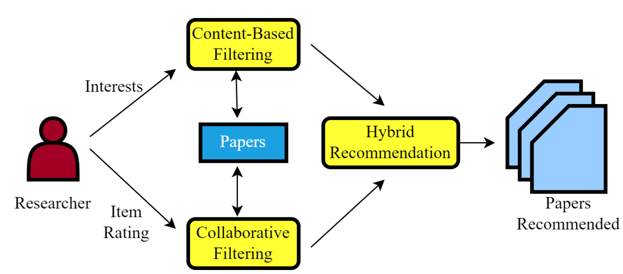

Fig. 5. Hybrid Filtering

II. Different Methods of Hybrid Recommendation Systems

The Semantic-Web Usage Mining proposal by Sharma et al.[3] asserts that personalized search based on long-term behavioral cues can be on par with or superior to the state-of-the-art in this field. Gandhi and Gandhi[10] employed collaborative filtering and association rule mining in order to guarantee strong scalability and strength of the recommendation system. Algorithms for collaborative filtering and association rule mining should also be parallelized in order to manage huge data. A hybrid approach using big data methods like association rule mining along with RS techniques like CF provides a comprehensive, durable recommendation system with high accuracy. There are three steps to the proposed work. After pre-processing, the unstructured data is transformed into a structured form, and a user-item matrix is created. From the experimental results, it was clearly observed that the hybrid model is performing well than the other methods in terms of using the information appropriately to enhance the performance in terms of scalability and strength.[10]
Tarus et al.[11] used "Ontology and Sequential Pattern Mining (SPM)"[11] and came up with a hybrid knowledge-based RS to suggest the learners with accurate e-learning resources. E-learning recommender systems' main goal is to steer learners toward accurate, suitable learning materials. This paper employs SPM for medicine prediction and ontology for context modeling. By utilizing the total user-item single-criterion, researchers are aiming to increase the recommender systems' accuracy. In this suggested study, ontology is used to model and express the domain information about learners and their learning resources. Whereas the learner's sequential learning patterns are discovered by the SPM algorithm and are illustrated in Fig 6. Filter the suggestions based on the learner's previous sequential learning patterns considering the list of top N recommendations and applying the GSP algorithm to it. The GSP algorithm is used to extract the sequence from weblogs. The experimental results show that in terms of performance and prediction accuracy, the hybrid algorithm "(CF+Onto+SPM)"[11] beats the traditional CF.

Wasid and Ali[12] in 2019 created a multi-criteria recommender system (MCRS) to express user preferences for several elements of the objects. This paper suggests a clustering technique using multi-criteria ratings. You may provide suggestions that are more accurate by utilizing the Mahalanobis distance approach to determine intra-cluster user similarities. In this approach, user preferences are derived using multi-criteria ratings, and different user groups are identified using similarity metrics like Pearson correlation, cosine similarity, or Euclidean distance. Users will be assigned to suitable clusters until convergence requirements are met, this procedure iterates. The Mahalanobis distance (MD) method is used to convert the dataset's variance-covariance matrix into its inverse, and it is also calculated for a single user. The distance between user-co-rated item ratings is calculated using the Mahalanobis distance algorithm. The similarity between various users within the same cluster after generating groups is computed in order to find the target user's top-N neighbors. The experiment's objective was satisfied by the performance statistic known as Mean Absolute Error (MAE), which is straightforward and accurate. Where it is stated that "n is the number of ratings, r_i is the item's actual rating, and p_i is the target user's anticipated rating on item i[12], we may compute the MAE.

$$MAE = \frac{\sum_{i=1}^{n} |p_i - r_i|}{n} \quad (1)$$

The proposed technique and the current methodology were evaluated in both clustering and non-clustering settings for Yahoo! Movies dataset. Table 1 and Fig. 7 compare the effectiveness of the Mahalabosis distance recommendation system (MDRS) and the current Pearson collaborative recommender (PCRS). As is evident from the findings, both in non-clustering and clustering situations, In aspects of the MAE of the System, MDRS fared much better than PCRS.

TABLE I
Clustering and Non-clustering MAE in Collaborative Recommenders

Method	(Clustering) MAE	(Non-Clustering) MAE
MDR	2.175	2.3094
PCRS	2.273	2.4577

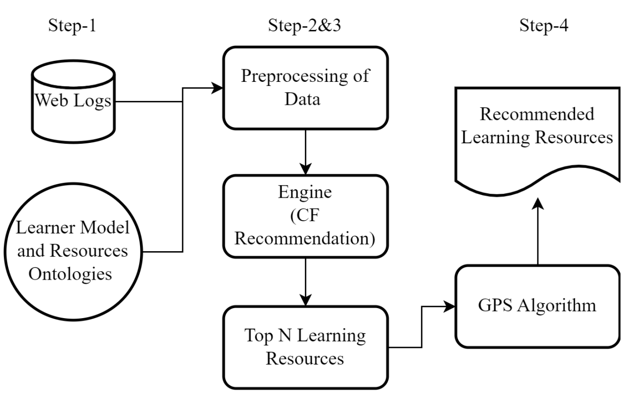

Fig. 6. Ontology and SPM-based Hybrid e-learning recommendation model

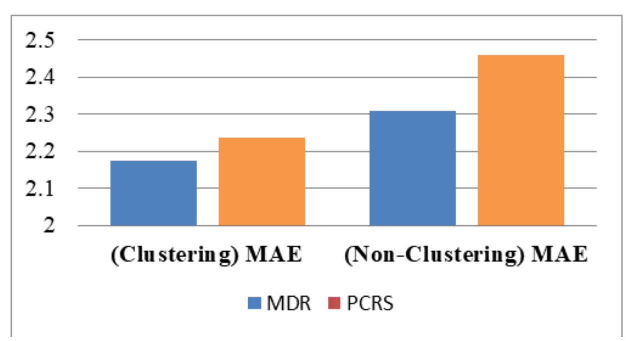

Fig. 7. Comparison of Collaborative recommenders based on clustering mechanisms

Sharma et al.[3] developed a unique strategy for customizing web searches using methodologies, applications, and possibilities of data mining and web mining. The major objective of the suggested approach is to investigate prior behavior to resolve ambiguity and deliver desirable solutions in response to perplexing requests. Semantic-Web Usage Mining's proposed approach comprises of four components, Query expansion, Ontology-based classification, Semantic clustering, and Statistical computation depicted in Fig.8:

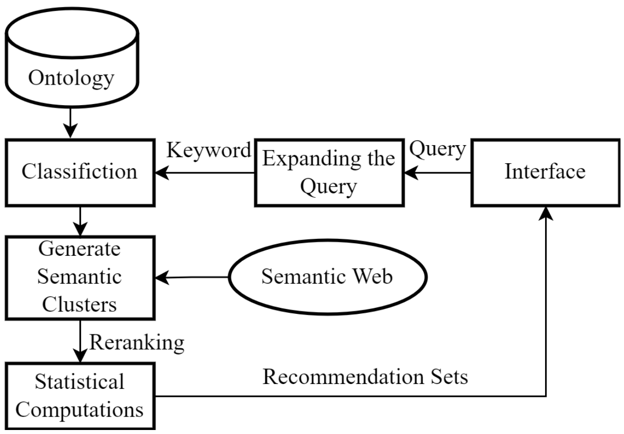

Fig. 8. Semantic-web usage mining

The first relevant document, or the number of significant results per query created by the four components named earlier, serves as a junction between search results and documents whose results are relevant to the user query. As a result, a successful e-commerce firm may be built by developers and other stakeholders.

Dong et al.[15] developed a hybrid model that simultaneously conducts deep user and item latent factor learning from side information and collaborative filtering from the rating matrix. The hybrid model successfully beats previous techniques, according to extensive experimental findings on three real-world datasets. It solves the cold start problem and data sparsity problem by extending the stacked denoising autoencoder to incorporate more side information into the inputs. This new learning model is known as an additional stacked denoising autoencoder (aSDEA). This strongly links collaborative filtering for the rating matrix and deep representation learning for the additional side information in the hybrid model.

Zhao[14] provides a framework for a Hadoop-based distributed and scalable recommender system combining hybrid recommendation techniques with the processing power and scalability of MapReduce, , addressing the issue of information overload in large-scale e-commerce. The structure of the Hadoop-based recommender system is achieved through a layered architecture. The system's versatility is ensured by the ability to configure the parameters and metrics of hybrid recommendation models. The benefits of Hadoop's distributed computing capabilities, mixed recommendation, as well as its scalable, adaptable, and diversified recommender system, may be used to solve the information overload problem in large e-commerce.

Rahim et al.[13] introduced the TrustASVD++ trust-based approach, which integrates user trust data in the Matrix Factorization context seen in Fig 9. It utilizes Pearson Correlation Coefficient to integrate trust metrics with user ratings for better suggestions (PCC). The ASVD++ approach handles the issues of sparse data and cold-start users in the first stage by using matrix factorization. Finding comparable users occurs in the second step, which makes use of rating and trust data. Using the similarity metric (PCC) the target user is then given recommendations based on other users who are similar to them. This method performs better than the CB-based method, which solely considers user and item similarity.

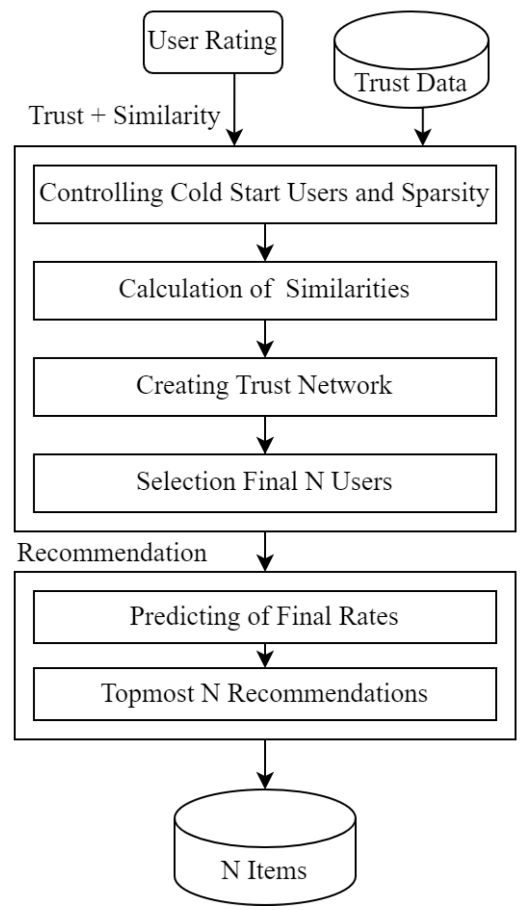

Fig. 9. Working Mechanism of TrustASVD++ trust-based Approach

Ramzan et al.[8] proposed an intelligent method that utilizes large-sized heterogeneous data to address the needs of potential customers and generate authentic suggestions. A hotel feature matrix is generated via an innovative Collaborative Filtering recommendation technique that employs opinion-based sentiment analysis and polarity recognition. The system functions in two separate ways. Each chosen data source's normalized hotel rankings and reviews are in numeric form. In order to mine reviews, natural language processing software is utilized to evaluate review data, then features are extracted using a hotel feature matrix. Additionally, the average polarity score is calculated for each of these retrieved characteristics using SentiWordNet. In order to create the weighted average polarity scores, the normalized rank score, voting score, and polarity score were combined. The fuzzy logic technique is then used to compute suggestions. The subject matter expert uses fuzzy sets to describe the hotel attributes and customer traits, while recommendations are given based on a formerly created hotel-feature-rating matrix. The proposed recommendation system is used to manage large datasets including user reviews, rankings, votes, and video views. A sentiment analysis approach is used by the recommender to provide consumers with precise and excellent suggestions.

Won et al.[6] designed a new hybrid CF model based on Doc2Vec that employs search terms and purchase history data from online shopping mall users. Doc2Vec is a Word2Vec extension that detects and transforms relationships between words in sentences. Doc2Vec uses an elementary ANN to detect comparable papers and then locate those papers in close proximity on multidimensional surfaces. Doc2Vec has two approaches: "Distributed Memory (DM) and Distributed Bag of Words (DBOW)".[6] The DM strategy, which is comparable to Word2Vec's CBOW (Continuous Bag of Words) method, combines the document's word vectors and document ID vector to predict the presence of additional words in the text. The DBOW technique ignores word order in the text and forecasts a document based only on its words. Fig 10 illustrates the proposed model. Experiments indicate that search terms data may successfully represent consumer preferences and contribute to the enhancement of traditional CF.

Yang et al.[9] created a hybrid RS that combines Collaborative and CB similarity models with Markov chains for the sequential suggestion "(called U2CMS)"[9]. The proposed model is divided into three components, modeling the CB filtering, Modeling User Preferences, and Modeling Sequential Patterns.U2CMS considers both sequential patterns and content information for determining proper connections between things. Using a higher-order Markov chain to replicate sequential patterns across a number of time steps and text from the content, it models the recommender system. Several studies were conducted on various Amazon datasets which are used to demonstrate the usefulness of the U2CMS in dealing with sparsity difficulties, diverse N-ordered Markov Chains, and properly finding similarities among the items. The results reveal that the U2CMS not only outperforms existing state-of-the-art RS but it also effectively tackles sparsity challenges effectively than other techniques.

Patro et al.[7] presented a unique methodological technique termed "Sparsity and Cold Start Aware Hybrid Recommended System (SCSHRS)" to reduce CSP and data sparsity in RS. The SCSHRS approach makes the following contributions, as indicated in Fig 11: Methods such as "SRCF (Sparsity Resolving Collaborative Filtering) and SRWCF (Sparsity

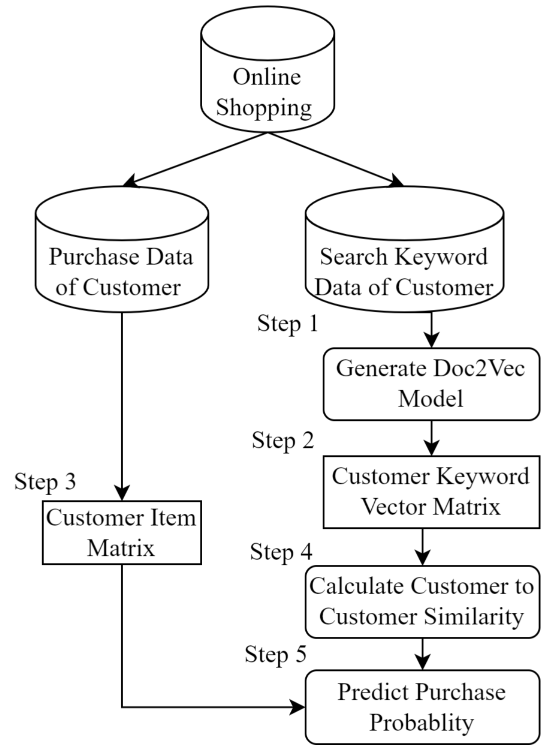

Fig. 10. Illustration of Hybrid collaborative filtering (CF) Using Doc2Vec Approach

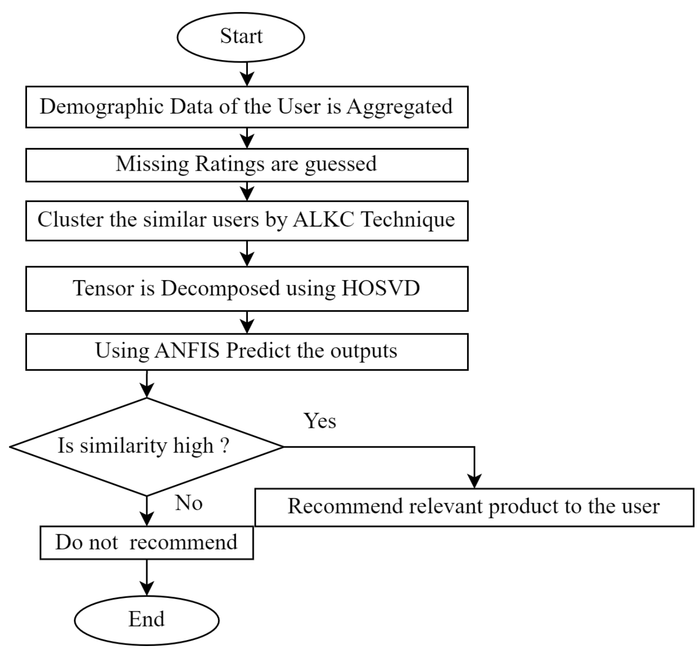

Fig. 11. SCSHRS approach

Resolving Weighted Collaborative Filtering)"[7] are used to forecast unavailable ratings. The clustering process uses Ant Lion Optimization (ALO), which helps determine the k-means starting point and increases the precision of suggestions. In addition, "HOSVD (Higher-Order Singular Value Decomposition) decomposes higher-order data to lower dimensions"[7]. This speeds up the calculation. In order to acquire information, the ANFIS technology is used. This approach does not rely on individual specialists, but rather on decreasing mistakes through a training process. This results in a speedy and accurate forecast. The investigation's findings indicate that when it comes to performance and inaccuracy, the SCSHRS approach is highly recommended.

III. Conclusion and Future Work

Various hybrid recommendation techniques have been discussed in this paper not only to highlight the strength of the techniques used but also to highlight the challenges associated with various types of hybridization strategies that are utilized to enhance performance. The experimental findings demonstrated that, in terms of accuracy and performance, the suggested hybrid algorithms performed better than those of other comparable algorithms, and also, the proposed systems helps in overcoming the data sparsity problems and cold-start problems. However, in order to further improve the performance and accuracy of recommendations, the focus of future work will be incorporating intelligent tools and technologies into the recommendation process from several fields, including data mining, machine learning, deep learning, and Genetic Algorithms.

References

[1] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. Mahboob Alam, and S. Luo, "A Review of Content-Based and Context-Based Recommendation Systems," Int. J. Emerg. Technol. Learn. IJET, vol. 16, no. 03, p. 274, Feb. 2021, doi: 10.3991/ijet.v16i03.18851.

[2] H. Papadakis, A. Papagrigoriou, C. Panagiotakis, E. Kosmas, and P. Fragopoulou, "Collaborative filtering recommender systems taxonomy," Knowl. Inf. Syst., vol. 64, no. 1, pp. 35–74, Jan. 2022, doi: 10.1007/s10115-021-01628-7.

[3] S. Sharma, S. Mahajan, and V. Rana, "A semantic framework for ecommerce search engine optimization," Int. J. Inf. Technol., vol. 11, no. 1, pp. 31–36, Mar. 2019, doi: 10.1007/s41870-018-0232-y.

[4] R. Sharma, S. Rani, and S. Tanwar, "Machine Learning Algorithms for building Recommender Systems," in 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, May 2019, pp. 785–790. doi: 10.1109/ICCS45141.2019.9065538.

[5] R. Chen, Q. Hua, Y.-S. Chang, B. Wang, L. Zhang, and X. Kong, "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, doi: 10.1109/ACCESS.2018.2877208.

[6] H.-R. Won, Y. Lee, J.-S. Shim, and H. Ahn, "A Hybrid Collaborative Filtering Model Using Customer Search Keyword Data for Product Recommendation," in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, Dec. 2019, pp. 1523–1526. doi: 10.1109/ICMLA.2019.00251.

[7] S. G. K. Patro, B. K. Mishra, S. K. Panda, R. Kumar, H. V. Long, and D. Taniar, "Cold start aware hybrid recommender system approach for E-commerce users," Soft Comput., Aug. 2022, doi: 10.1007/s00500-022-07378-0.

[8] B. Ramzan et al., "An Intelligent Data Analysis for Recommendation Systems Using Machine Learning," Sci. Program., vol. 2019, pp. 1–20, Oct. 2019, doi: 10.1155/2019/5941096.

[9] Y. Yang, H.-J. Jang, and B. Kim, "A Hybrid Recommender System for Sequential Recommendation: Combining Similarity Models With Markov Chains," IEEE Access, vol. 8, pp. 190136–190146, 2020, doi: 10.1109/ACCESS.2020.3027380.

[10] S. Gandhi and M. Gandhi, "Hybrid Recommendation System with Collaborative Filtering and Association Rule Mining Using Big Data," in 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, Apr. 2018, pp. 1–5. doi: 10.1109/I2CT.2018.8529683.

[11] J. K. Tarus, Z. Niu, and D. Kalui, "A hybrid recommender system for e-learning based on context awareness and sequential pattern mining," Soft Comput., vol. 22, no. 8, pp. 2449–2461, Apr. 2018, doi: 10.1007/s00500-017-2720-6.

[12] M. Wasid and R. Ali, "An Improved Recommender System based on Multi-criteria Clustering Approach," Procedia Comput. Sci., vol. 131, pp. 93–101, 2018, doi: 10.1016/j.procs.2018.04.190

[13] A. Rahim, M. Y. Durrani, S. Gillani, Z. Ali, N. U. Hasan, and M. Kim, "An efficient recommender system algorithm using trust data," J. Supercomput., vol. 78, no. 3, pp. 3184–3204, Feb. 2022, doi: 10.1007/s11227-021-03991-2.

[14] X. Zhao, "A Study on E-commerce Recommender System Based on Big Data," in 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, Apr. 2019, pp. 222–226. doi: 10.1109/ICCCBDA.2019.8725694.

[15] X. Dong, L. Yu, Z. Wu, Y. Sun, L. Yuan, and F. Zhang, "A Hybrid Collaborative Filtering Model with Deep Structure for Recommender Systems," Proc. AAAI Conf. Artif. Intell., vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.10747.

[16] M. C. Urdaneta-Ponte, A. Mendez-Zorrilla, and I. Oleagordia-Ruiz, "Recommendation Systems for Education: Systematic Review," Electronics, vol. 10, no. 14, p. 1611, Jul. 2021, doi: 10.3390/electronics10141611

[17] S. Sharma, V. Rana, and M. Malhotra, "Automatic recommendation system based on hybrid filtering algorithm," Educ. Inf. Technol., vol. 27, no. 2, pp. 1523–1538, Mar. 2022, doi: 10.1007/s10639-021-10643-8.

[18] M. Marcuzzo, A. Zangari, A. Albarelli, and A. Gasparetto, "Recommendation Systems: An Insight Into Current Development and Future Research Challenges," IEEE Access, vol. 10, pp. 86578–86623, 2022, doi: 10.1109/ACCESS.2022.3194536.

[19] F. Fkih, "Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison," J. King Saud Univ. - Comput. Inf. Sci., p. S1319157821002652, Sep. 2021, doi: 10.1016/j.jksuci.2021.09.014.

[20] B. Walek and V. Fojtik, "A hybrid recommender system for recommending relevant movies using an expert system," Expert Syst. Appl., vol. 158, p. 113452, Nov. 2020, doi: 10.1016/j.eswa.2020.113452.