International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Mixed Recommendation Algorithm Based on Content, Demographic and Collaborative Filtering

Isha Nadkar[1], Dr. Jaydeep B. Patil[2].

[1]*Student, Department of Information Technology, AISSMS Institute of Information Technology, Pune, Maharashtra, India*
[2]*Assistant Professor, Department of Information Technology, AISSMS Institute of Information Technology, Pune, Maharashtra, India*

Abstract - *This paper investigates, and analyses recommendation technologies based on content filtering and user collaborative filtering, demographic filtering, fuzzy logic and introduces a hybrid recommendation algorithm. This method not only takes advantage of the power of content filtering and demographic filtering, but it can also do similarity matching filtering for all items, especially when the items have not been reviewed by any user, enabling them to be filtered out and recommended to users, avoiding the early level problem. At the same time, this methodology takes advantage of collaborative filtering's features. The user rating data matrix of collaborative filtering prediction becomes reasonably dense when the number of users and evaluation levels are high, which minimizes the sparsity of the matrix and enhances collaborative filtering precision. The performance of the system will be significantly improved as a result of the integration of the two. A hybrid algorithm was established based on the improved hybrid content-based and collaborative filtering algorithm. A user feature rating matrix was developed to replace the traditional user-item rating matrix by amalgamating user rating with item features. On the user set, K-means clustering was performed, and recommendations were given. The enhanced algorithm can address the problem of data sparsity that classic collaborative filtering algorithms possess. Concurrently, it can forecast users who could be interested in new projects based on the matching of project and user attributes scoring matrix and build a push list for new projects, thereby solving the problem of "cold start" for new projects. The experimental results suggest that the modified algorithm in this study plays a substantial role in tackling the speed bottleneck problems of data sparsity, cold start, and online recommendation, as well as ensuring a superior recommendation rate.*

Key Words: Recommendation Systems, Recommendation Algorithms, Content-based filtering, Collaborative -filtering, Demographic-filtering, Hybrid Algorithm, Mixed Algorithm.

1. INTRODUCTION

With the growth of YouTube, Amazon, Netflix, and other similar web services over the last few decades, recommender systems have become increasingly important in our lives. Recommender systems are becoming ubiquitous in our daily online trips, from e-commerce (suggest articles to buyers that may be of interest) to online advertising (suggest the suitable contents to users based on their preferences).

A recommendation system mines the mass information for resources that users might be interested in or require based on their interest criteria and generates recommendations. It is acknowledged as one of the most viable methods for dealing with the information explosion. In principle, the recommendation system quantifies the likes of some items that the user has never explored by analysing the resources chosen by the user and presenting the user with the products with the strongest likes among the projected outcomes.[3][5][9][11][2]

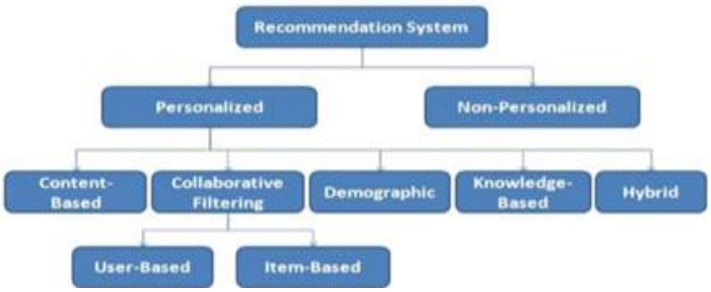

Fig -1: Classification of Recommendation Systems

1.1 Personalized Recommendation Systems

1. **Content-Based:** A content-based recommender uses information provided by the user, either intentionally (reviews) or unintentionally (search queries) (clicking on a link). A user profile is created based on this information, and it is then utilized to provide recommendations to the user. The engine becomes increasingly accurate as the user offers more inputs or acts on the recommendations.[4][12][7][2]

2. **Collaborative Filtering:** We tend to locate similar users and recommend what they like in Collaborative Filtering. We don't use the item's features to recommend it in this sort of recommendation system; instead, we group users

into clusters of similar categories and recommend each user based on the preferences of their cluster.[4][12][7][2]

3. **Demographic Filtering:** The Demographic Recommender system makes recommendations based on the demographic characteristics of the user. It categorizes people based on their characteristics and suggests products based on their demographic information. It is simple to accomplish and does not require user ratings, unlike collaborative filtering and content-based recommender systems.[10][2]

4. **Knowledge-Based:** Knowledge-based recommender systems are a type of recommender system that is built on direct knowledge of the item selection, user preferences, and suggestion criteria.[2]

5. **Hybrid:** Mixed approach combining 2 or more recommendation generation methods. Their main target is to eliminate the drawbacks of the individual ones.[12][8][2]

1.2 Non-Personalized Recommendation Systems

A personalised recommendation system proposes products to a user based on their previous purchase history, whereas a non-personalized recommender system shows things that are popular among the broader public at the time. A recommendation system assists a website in building its first impression of having things that may be of interest to the user by proposing a product that the customer may potentially buy. Non-personalized systems come in a variety of shapes and sizes, depending on the needs of the customer.[2]

2. LITERATURE REVIEW

1. **A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems:** This paper provides a summary analysis of literature studies in accordance with the film recommendation system. In this analysis, behavior of the algorithms in multiple contexts is compared and not just under the most favorable conditions, with an optimistic comparison of different collaborative filtering algorithms. Similarly, using a clearly defined approach would make it easier in the future to compare other approaches to solve one of the major problems when evaluating collaborative filtering algorithms. It is interested that new algorithms and other datasets are used in future studies in various fields. It can also be especially important to apply the trend-based algorithm to contexts like knowledge recuperation on the web. [4]

2. **Hybrid Recommender Systems: A Systematic Literature Review:** This systematic literature review presents the state of the art in hybrid recommender systems of the last decade. It is the first quantitative review work completely focused on hybrid recommenders. We address the most relevant problems considered and present the associated data mining and recommendation techniques used to overcome them. We also explore the hybridization classes each hybrid recommender belongs to, the application domains, the evaluation process and proposed future research directions. Based on our findings, most of the studies combine collaborative filtering with another technique often in a weighted way. Also, cold-start and data sparsity are the two traditional and top problems being addressed in 23 and 22 studies each, while movies and movie datasets are still widely used by most of the authors. As most of the studies are evaluated by comparisons with similar methods using accuracy metrics, providing more credible and user-oriented evaluations remains a typical challenge. Besides this, newer challenges were also identified such as responding to the variation of user context, evolving user tastes or providing cross-domain recommendations.[8]

3. **Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation**: This paper explores, and studies recommendation technologies based on content filtering and user collaborative filtering and proposes a hybrid recommendation algorithm based on content and user collaborative filtering. This method not only makes use of the advantages of content filtering but also can carry out similarity matching filtering for all items, especially when the items are not evaluated by any user, which can be filtered out and recommended to users, thus avoiding the problem of early level. Based on the improved collaborative filtering algorithm, a hybrid algorithm based on content and improved collaborative filtering was proposed. By combining user rating with item features, a user feature rating matrix was established to replace the traditional user-item rating matrix. K-means clustering was performed on the user set and recommendations were made.[1]

4. **DECORS:** A Simple and Efficient Demographic Collaborative Recommender System for Movie Recommendation DECORS is based on collaborative filtering which initially partitions the users based on demographic attributes, then using k-means clustering algorithm clusters the partitioned users based on user rating matrix. It reduces the expensive computations to identify similar users in order to

predict movies when compared with traditional collaborative filtering approach. The proposed system also sorts the movies in the increasing order of user's preferences. The proposed framework is assessed by using the performance measurement MAE (Mean Absolute Error). The results proved that proposed system is more efficient when compared against traditional methods.[10]

5. **Course recommendation system using fuzzy logic approach:** This paper proposed to develop a course recommendation system using fuzzy logic approach. The development methodology of this system involves several phases include requirements planning, user design and construction for prototyping, testing and cutover. This study used the fuzzy rules technique in order to calculate each associated student's skill and interest level based on Mamdani fuzzy inference system method. Then, the rules will generate outcome which recommend the suitable course path and provide the details to a user based on their course test. The result shows the functionality of this system has been achieved and works well. This study is significantly helping the students to choose their course based on the interest and skill.[6]

3. METHODOLOGY

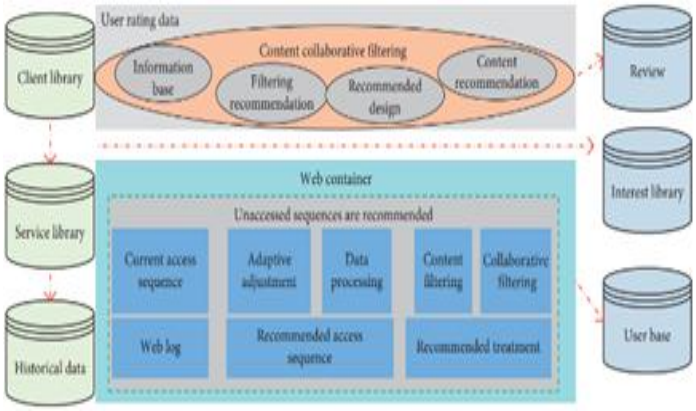

Fig -2: Architectural framework for the hybrid mode recommendation algorithm based on content and collaborative filtering.

The comprehensive recommendation is split into two main modules: content filtering recommendation and collaborative filtering recommendation, both of which are hidden from users.[1]

Following is the algorithm for the given methodology: -

1. Initially, the user's interest is extracted from the user's purchasing history data, and the topic vector and feature vector are pre-processed by the network log, and the data processing is established based on the content filtering recommendation module.
2. Then, leveraging user preference features, user rating data, and current access sequence data, a collaborative filtering-based recommendation module is developed to extract the user's nearest neighbours and the current access sequence's nearest neighbours (item).
3. Then, to establish the recommended top visit sequence, it incorporates two recommendation weighted sum calculation modules for the similarity calculation model of mixed recommendation (i.e., recommendation processing).
4. To achieve the optimum recommendation data, the web server proposes the sequence to the user and retrieves the recommendation sequence on the user, the dynamic modification of the suggested framework, and the idle speed value of the feedback information. The words with the highest mutual information amount with the related text set Rel (Q) are called optimal feature items. The logarithmic mutual information amount between the words and the related text set is calculated using the equation below:

$$\mathbf{logNI\ (\alpha, rel(Q)) = log\ (Q\ (\alpha| \in rel(Q)/Q(\alpha)))}$$

The cosine similarity between user preference document and project document is:

$$\mathbf{sin\ (\beta, ej) = cos\ (ej, ec) = \Sigma i\alpha i\varepsilon ij/\alpha i\varepsilon i}$$

The higher the calculated similarity, the more preference the users have for this feature.

5. Clustering is done primarily on how consistent user feature vectors are, so that users with shared interests can be clustered together for simplified processing. Meanwhile, for new product information documents, evaluating their categories can yield a list of suggested users. Since it is anticipated that user sets are categorized manually, the recommender system clustering approach can be used.
6. The correlation between the content about the product introduction in the product information database and the model vector of a user's interest subject can be computed once the initial recommendation model has been structured and the threshold has been set. It is classified associated with the user's interests if the similarity value is more than or equal to fujian. You can make a recommendation to the user via the Web server and then validate whether the recommendation is accurate.

4. PROPOSED METHODOLOGY

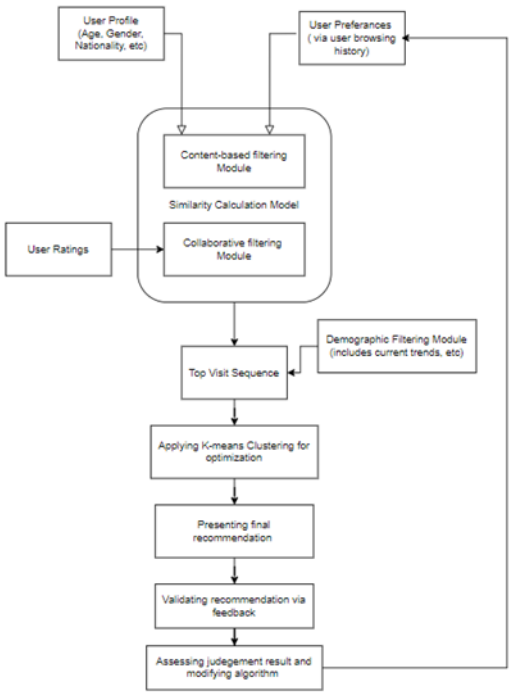

Fig -3: Proposed Methodology

Figure above illustrates the flow of the proposed methodology which is upgraded variant of the above-mentioned approach by factoring in the demographic of the user profile and current market trends along with fuzzy logic in the existing methodology.[1][10][6]

Following the is the approach for the proposed methodology:

1. **Content-based filtering Module:** A user profile will be developed based user attributes such as:

   • Age

   • Gender

   • Economic Background

   • Nationality

   • Ethnicity, etc.

   This profile will be integrated with the existing Content-based filtering module. Therefore, the revised module will not only be based on user preferences (extracted via user browsing history) but will also include various user attributes.

2. **Collaborative- filtering Module**: This module will be developed by considering user preference features, user rating data, and current access sequence data. The module will resemble to the Collaborative -filtering Module proposed in the Original Methodology.

3. **Demographic filtering Module:** This module will be constructed by factoring in attributes such as region based current trends, nation specific sociocultural occasions, medical trends, market trends etc.

4. **Similarity Calculation Model:** For the similarity calculation model of mixed recommendation, two main recommendation weighted sum calculation modules i.e., Content-based Module and Collaborative-filtering Module are encompassed.

5. **Top Visit Sequence:** The similarity calculation model on further integration with the Demographic filtering module will be used to construct the suggested top visit sequence (i.e., recommendation processing).

6. **Applying K-means clustering:** On applying K-means clustering, the target user's clustering cluster is determined based on the correlation between the target user and each clustering center, in order to locate the cluster's nearest neighbor. Then, depending on the preferences of the project's closest neighbors, we can estimate the target users' interests and, ultimately, generate a recommendation.

7. The final recommendation generated will be presented to the user via web servers. Validation in form of feedback form will be acquired from the consumer. The validation will be further assessed using fuzzy logic.

8. According to the judgement result, the system will dynamically alter the model vector or adjust the proximity value to obtain optimum performance.

4. ANALYTICAL WORK

4.1. Comparison of Original Methodology with Proposed Methodology

Sr. No.	Original Methodology	Proposed Methodology
1.	Uses content and collaborative filtering approach.	Integrates demographic filtering and fuzzy logic to the existing approach

2.	Heavily based on user manual inputs such as ratings provided by user; user preferences accessed via browsing history.	The proposed approach not only focuses on user inputs but also factors in other prospects such user attributes, current trends etc.
3.	Simply validates the recommendation from user but there are not provisions for usage of this validation for updating the algorithm to incorporate the user's latest preferences.	The proposed method not only gets feedback from the user for the provided recommendations but also assesses using fuzzy logic and constant keeps updating the user preferences.

4.2. Advantages and Disadvantages of the Proposed Methodology

Advantages	**Disadvantages**
System provides highly personalized and researched recommendations.	Accurate recommendations are not guaranteed.
System is not entirely dependent on user preferences and user-product ratings.	Increase in data or large dataset may affect the quality of recommendations.
The proposed algorithm handles the issue of cold start i.e., therefore performance of newly launched products isn't hampered	As fuzzy logic works on precise as well as imprecise data so most of the time accuracy is compromised.
Fuzzy logic can work with imprecise, distorted judgement results while assessing feedback.	As this approach takes in account multiple factors, the time complexity of the algorithm is quite high.
Algorithm keeps updating user interests within the system by using the feedback	

5. CONCLUSION

A mix recommendation solution based on several recommendation generating approaches was developed, combining the benefits and drawbacks of content, demographic, and collaborative filtering. This recommendation technology's workflow, user characteristic description, data processing method, and recommendation approach were all thoroughly investigated. This method not only takes advantage of the benefits of content filtering, but it can also do similarity matching filtering for all items, particularly when the items have not been reviewed by any user, and can be filtered out and recommended to users, avoiding the early level problem. Along with this, the algorithm keeps dynamically updating with change in user interests. This recommendation solution is expected to provide relevant and accurate results.[2]

6. REFERANCES

[1] Lianhuan Li, Zheng Zhang, and Shaoda Zhang. "Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation". In: Scientific Programming 2021 (2021).

[2] Zeshan Fayyaz et al. "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities". In: applied sciences 10.21 (2020), p. 7748.

[3] Jing Li and Zhou Ye. "Course recommendations in online education based on collaborative filtering recommendation algorithm". In: Complexity 2020 (2020).

[4] Sri Hari Nallamala et al. "A Brief Analysis of Collaborative and Content Base Filtering Algorithms used in Recommender Systems". In: IOP Conference Series: Materials Science and Engineering. Vol. 981. 2. IOP Publishing. 2020,p. 022008.

[5] Bo Song, Yue Gao, and Xiao-Mei Li. "Research on collaborative filtering recommendation algorithm based on mahout and user model". In: Journal of Physics: Conference Series. Vol. 1437. 1. IOP Publishing. 2020, p. 012095.

[6] Mohd Suffian Sulaiman et al. "Course recommendation system using fuzzy logic approach". In: Indonesian Journal of Electrical Engineering and Computer Science 17.1 (2020), pp. 365–371.

[7] Parul Aggarwal, Vishal Tomar, and Aditya Kathuria. "Comparing content based and collaborative filtering in recommender systems". In: International Journal of New Technology and Research 3.4 (2017), p. 263309.

[8] Erion C¸ ano and Maurizio Morisio. "Hybrid recommender systems: A systematic literature review". In: Intelligent Data Analysis 21.6 (2017), pp. 1487–1524.

[9] Tom´aˇs Horv´ath and Andr´e CPLF de Carvalho. "Evolutionary computing in recommender systems: a review of recent research". In: Natural Computing 16.3(2017), pp. 441–462.

[10] M Sridevi and R Rajeswara Rao. "Decors: A simple and efficient demographic collaborative recommender system for movie recommendation". In: Advances in Computational Sciences and Technology 10.7 (2017), pp. 1969–1979.

[11] Rabi Narayan Behera and Sujata Dash. "A particle swarm optimization-based hybrid recommendation system". In: International Journal of Knowledge Discovery in Bioinformatics (IJKDB) 6.2 (2016), pp. 1–10.

[12] Poonam B Thorat, RM Goudar, and Sunita Barve. "Survey on collaborative filtering, content-based filtering and hybrid recommendation system". In: International Journal of Computer Applications 110.4 (2015), pp. 31–36