ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management, Vol-5, Issue-6, 2021.

Hybrid Recommender System based on Collaborative Filtering and Eclat Rule Mining for Recommending Movies

Rohit Dharmik[1], Dr. S. K. Pandey[2]

Research Scholar[1], CSE Department[1,2]
VNS Group of Institute Bhopal (M.P.)[1,2]

Abstract-: *An intelligent system has recently been developed and is not a passive system for the right service. The latest system will respond to the user's request and recommend him. It is used for personalized recommendation, and the most common ways are content-based and collaborative filtering (CF). To address the issue of sparsity and collaborative filtering, we proposed a prediction technique based on a recommendation system (RS) and a collaborative filtering or rule mining (RM) strategy. To validate the prediction approach, we used consumer rating data on movies. To estimate the user's rating, utilize this Movie Lens 1M dataset. The accuracy of the method's Root Mean Square Error (RMSE) and Mean Square Error (MSE) is determined by comparing it to the real value of each movie. The present system is compared in addition to the projected RMSE and MSE.. In addition to the expected RMSE & MSE, the existing system is compared. This work also gives a comparative study of a variety of similarity measures, including the Cosine, Cityblock, Chevyshev, Braycurtis, and Spearman correlation. Rule mining with the Eclat rule is completed for the ideal recommendation on a rating matrix. Recall, Precision, & F1 score measure is used to assess recommendation accuracy. From the experimental results, we can show that Cityblock and Chevyshev measure based Eclat rule mining achieved better precision and F1 Score in comparison to other measures.*

Keywords:- Recommender System, Collaborative Filtering, Association Rule Mining, Eclat Algorithm, Similarity Measures.

Introduction

An informal community is a social structure comprised of people called hubs, which are tied by at least one explicit kinds of interdependence, for example, kinship, family relationship, normal interest, hates, convictions. Interpersonal organization investigation analyses the structure of social connections in gathering to reveal a casual association between individuals. Informal organization examination depends on the presumption of the significance of connections among associating units. It shows how they are associated through various social familiarizes going from easy-going colleagues to close recognizable bonds. The perspective of casual association wraps up ideas, models & applications that are imparted about social thoughts and cycles. Close by generating revenue, extended use of association examination, an arrangement on the fundamental association perspective of central norms has come[1][2].

The rapid emergence of large scholarly data is now mostly digitally captured and archived from around the world. Archival material is now actually digitized and distributed provided for free or via a fee online to people. Such a situation aims problem of data overloading, especially in academia. For example, if a researcher wants to find any article related to his research, he is writing its need to choose and select papers from a huge quantity of data. The process of filtering is generally tiresome and time-consuming; therefore, so many researchers have focused their attention on the ease of the existing process system and make it more useful by providing better recommendations to the users [3][4].

Rule mining is widely used for exploring hidden similarities or connections between certain objects. In the discovery process, it includes each transaction in the database. It also indicates a set of strategies for respective development that can be followed and overlooked in the field. The association of fields with enormous data is mainly explored by RM. Many experiments on association law have been performed and an appropriate approach has been shown. RM is a data mining activity that recognizes interesting relationships among database variables. It aims at finding strong rules found in datasets using numerous measures. The rule can be described as a relationship of trust and support between items. [5][6].

The Recommendation System is in the daily life of people who rely on information to determine their interests. Sub-class information filtering is recommended as a prediction for a preference of items used by or for users. While several methods have been developed in the past, the search persists and it is also used in many apps, which tailor recommendations and resolve the overload of information. A recommendation system is a sharp system that gives users an idea of items that are amazon.com, movies in movies, music in last. fm[7]. The three major recommendation systems divisions are:

1.1 Content-based RS

It is used in one of the common RS. Concerning the product content, product attributes, product description, etc. the recommendation is done in this recommendation system. Describe the product attributes in the content-based recommendation system. In this method, there is an overview of the users'tastes and dislikes, the result of which is used for recommendation. information retrieval or information filtering are the main tasks in this recommendation framework. The compilation of information is a way to obtain specific information with aid of indexing & metadata. Data filtering is a technique that removes both unwanted, irrelevant, and duplicated data to increase process effectiveness. These different algorithms are cluster analyzes, decision tree algorithms, bayesian classifiers, artificial neural networks.

1.2 Collaborative based RS

The recommendations are determined by taking into account the behavior of users, user preferences, in a shared recommendations system. Without giving much attention to attributes and material, the system will accurately recommend complex items. In certain cases, associated with the content-based recommendation method, the recommendation system has been successful. Implicit data collection and explicit data collection are the key techniques involved. The implicit collection of data involves an analysis of the loves and dislikes of the user towards the item, an evaluation of how much he might have looked for a certain item, and an analysis of his interest in the object. Explicitly gathering data includes asking the user to score the object, alerting the items he has already taken, etc. numerous algorithms in this section are the K-nearest neighbor, algorithms for correlation, factorization & approximation algorithms.

1.3 Hybrid RS

This system combines a content-based system of recommendations with a collaborative recommendation system. In this RS, the recommendation is based on comparative performance, obtained by constantly tracking and analyzing user behaviors. It is more efficient in some cases than content-based recommendation systems and collaborative RS.

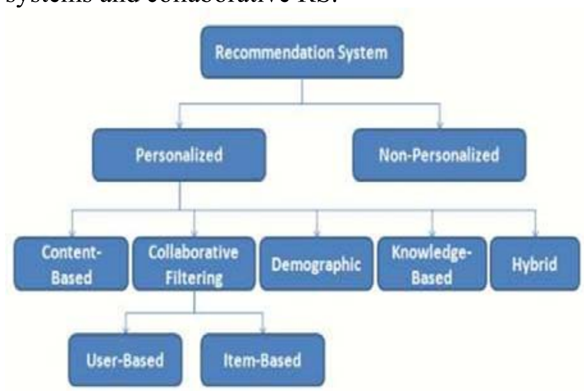

Figure 1: Types of RS.

Challenges & Issues of RS

At the early stages of a recommendation method, the issue of the cold start is clear. Ontologies are a validated method for the expansion of expertise where less evidence is available. When there is no content knowledge, then every recommender mechanism has the Cold Start problem, content-based filtering will be poorly performed and collaborative filtering will also be badly affected. If there is no content-based information recognizable and no history in the system, the hybrid approach often allows quite random recommendations.

1.4 Trust

The voices of those with a brief past might not be as important as the voices of those with a rich history. The trust issue arises about a customer's assessments. By sharing preferences with users, the issue can be solved.

1.5 Scalability

The system needs more tools to manage information or form recommendations for the increasing numbers of users and items. To identify users with like tastes and similar descriptions, the majority of services are consumed. The combination of different filter types and physical enhancement of systems also solves this problem. Many computations may also be carried out offline to accelerate online recommendations.

1.6 Sparsity

People purchase a few items from the vast number of products in e-commerce shops. Almost every user rated just a few items along with few ratings for the most common items. It is very difficult for a user to determine his or their choice if there are a few things available and they may have an association with the wrong neighborhood. Sparsity is the issue of the lack of consumer knowledge.

1.7 Privacy

The most critical issue is privacy. Many online shops maintain efficient privacy protection with the use of specialized algorithms and systems for their customers. To offer the latest and best recommendations, the device has to receive as much data on the user as possible, including demographic data and data on the user's position. The reliability, protection, and confidentiality of the information presented arise. CF approach has been proposed to address difficulties such as scalability and to increase the scalability of collaborative filtering algorithms and to reduce sparse data sets in the recommended system.[8][9].

The Movie Recommendation System offers a method to allow users with common interests to be classified. In concept, it is intended to search for material that is of interest to a person. The recommendation system also requires a variety of variables in creating individual lists for each user/individual of useful and interesting content. Artificial Intelligence-driven algorithms are recommendation systems that skip over all likely options & create a modified list of personally interesting items. These outcomes are based on their background, search/browse history, what other people are searching towards, and how likely they are to see these movies. These results are often based on their profile. This is done by using collaborative filtering based on items System.

The rest of thesis is organized as follows:-

Section 2:- outlines the related research background. And from studying the literature surveys.

Section 3:- presents the proposed algorithm and methodology.

Section 4:- describe the experiment and results.

Section 5:- conclusion of the paper.

II. Related Work

H. Ito et al. [10] Implement a collaborative filtering recommendation approach that is based on an association analysis that is methods of DM. They strive to enhance serendipity by using evaluation information that is distinct from a target customer to ensure precision. We also prove that precision and serendipity can be tailored to the proposed approach by a parameter. This work contrasts the system suggested with a traditional approach to accuracy and serendipity efficiency.

A. S. Tewari et al. [11] Recommendation systems are regularly used to warn buyers of products. Now websites that sell books online compete with each other on many occasions. The recommended method is one of the best methods to raise revenue and retain buyers. Books of consumer interest should be recommended using the book recommendation process. This paper introduces a book recommendation method based on the functionality, coordination, and management of combination rule filtering.
M. K. Kharita et al. [12] Propose a highly optimal and fast approach to recommendation systems. Although the principal aspect of the feature set is movie ranking alone, the derived method is very optimized and it is also possible to perform real-time analysis. The recommendation model is very complex and can regenerate new recommendations in real-time depending on the changes made by the subscribers. While the model's accuracy is somewhat poor relative to modern recommendations, our study stands out from the rest in terms of real-time recommendations. The work, item-based movie recommendation system, that uses movie rating as its function, has been identified. It utilizes a modified matrix of cosine similarity to find appropriate movies. RMSE that they get in the experiment is 1.01. They give a rating out of 5. The accuracy of item dependent CF algorithm is therefore 79.72 percent.

L. Uyangoda et al. [13] Authors use linkage of user feature-scores found from user-specific interaction by ratings in this proposed methodology to refine the input prediction algorithm parameters used in the recommended method to maximize prediction accuracy with fewer past data. In this case, the cold start problem solves a major drawback by evaluating the base CF algorithm with an increase of 8.4 percent. The system is generated and evaluated using the 'MovieLens 100k dataset.' The proposed method can also be applied to another field.

Z. Wang et al. [14] A new clustering algorithm is suggested in this work, together with a suggestion for collaborative filtering to increase recommendation efficiency. The suggested algorithm uses density peak clustering as its primary focus of a K-mean algorithm, which addresses the drawback that the K-means algorithm can effectively be implemented into local optimization, in advance by defining clustering center. The clustering algorithm is utilized to separate user also then CF recommendation is performed as per category to which target user belongs, that offers users with more "personalized" recommendation. New findings reveal that the suggested clustering algorithm increases the precision of the clustering relative to K-means. The Group Lens-MovieLens dataset also verifies the clustering algorithm along with the collaborative filtering recommendation, and results demonstrate that the algorithm is more efficient relative to the conventional CF recommendation algorithm.

F. Fessahaye et al. [15] Focuses on developing music recommendation mechanisms although several different platforms & domains, comprising Netflix (movies), Amazon (shopping), Youtube (videos), and so on can be added to the proposed solution. Present structures lack sufficient efficiency again with the introduction of variables. In the form of hybrid, collaborative filtering & CBF, our Tunes Recommendation System (T-RECSYS) offers real-time predictive feedback into a detailed classification

M. Gupta et al. [16] The goal of this work is to enhance the detailed and efficient filtering of a regular process. Though different approaches are used for applying RS, content-based filtering is the easiest approach. That takes user input, rechecks his/her history/past behavior, also recommends a list of the same movies. K-Nearest Neighbor (KNN) algorithms & CF are utilized in this work to show usefulness, especially to increase the accuracy of the findings relative to CBF. This solution is depending upon the cosine similarity of a KNN using CF to eliminate the limitations of the content-based filtering process. While it is chosen to use Euclidean distance, Cosine similarity is applied as the precision of cosine angle & movie equidistance are approximately the same.

Guo, J. et al. [17] Introduces a recommendation technique to comply with this restriction. Extended degree classification measures are firstly suggested to allocate items to further fine-grained classes. Later item seamlessness measurement is used to rapidly determine similarities among items in a similar class, which dramatically minimizes the runtime of similarity calculation. A correlation between the items based on Hellinger Distance (HD) is provided to measure item similarity regarding the distribution of rating probabilities. Also, the sigmoid function is studied in HD similarity to highlight the significance of co-rated items & efficiently differentiate b/w pair of items. Findings of tests on 2 benchmark datasets demonstrate that the suggested similarity approach by classification criteria has improved efficiency in both efficiency & accuracy related to additional approaches. Likewise, findings confirm the efficiency of proposed classification criteria, particularly runtime of item-based CF technique is decreased by a minimum 61 percent whilst preserving moderately stable or higher accuracy.

G. Suganeshwari and S. P. S. Ibrahim [18] present the ARM imputation method to enhance the top-N prediction performance of CF recommendation. The suggested approach uses the ARM strategy to classify and impute the unfavorable items of each user with low amounts. The solution suggested would not only address the problem of sparsity but will also boost the consistency of the decision by removing unfavorable items in Top-N predictions. Collaborative approaches are possible and can respond easily to the approach suggested, as the agnostic method. Experimental findings reveal that the proposed technique increases the precision of standard recommendation methods by 2 times on average and improves existing imputation-based methods substantially. Z. Chen et al. [19] introduce a new novel recommended algorithm that depends upon CF for course recommender to support student's decision. Enhanced cosine similarity is utilized, as per students' courseselection records history, & good accuracy is achieved in the recommendation task that sees users' requirements in this algorithm. Furthermore, both text vector & user behavior record are utilized to enhance the computation of course similarity. This work estimates 2022 students' 18457 records & 309 courses' real data. The experimental outcome demonstrates that the algorithm has a good outcome on recall rate, F1-score index & accuracy.

III. Proposed Methodology

In this proposal, we propose a new hybrid recommendation model system for movie recommendation using CF & Rule mining method. Firstly, collect Movies data from GroupLens Research. Then this dataset has been preprocessed to remove those movies rating is not high and users below than 20. We normally operate on extremely sparse matrices in advising structures because the item universe is very large and a single user generally deals with a very small portion of the item universe. This makes it incredibly sparse to represent users (as rows) & items (as columns) in a matrix, with several zero values. The rating matrix is then made by user-item rating. The rating matrix is used in Item-based CF for similitude measurement. The item similarity matrix is applied to construct the item-item matrix. We utilize 5 distinct similarity measures i.e. Cityblock, Cosine, Braycurtis, Chevyshev & Spearman correlation measures to achieve item-item similarity. In the same 5 similarity methods, the User-User similarity matrix is intended as well.

Proposed Algorithm

Input: MovieLens, 1M dataset.

Output: Accurate hybrid recommendation model for movie rating prediction.

Step 1. Collect the dataset of MovieLens, 1M data from public repositories.

Step 2. Removed least user (at least 20 users) viewed movies during preprocessing.

Step 3. Then, a rating matrix is produced by user-item rating and Find similarity based on user_id, movie_id, and rating.

Step 4. Calculate the sparse matrix for both user and movie using cosine measure and find error rate in terms of RMSE and MSE. Similarly, sparse matrix using remaining Cityblock, Chebyshev, Braycurtis, and Spearman correlation similarity measures.
Step 5. Once the sparse matrix has been generated, and the prediction matrix is measured for all similarity measures by applying the formula
Step 6. After implementing CF, the next step is producing frequently recommended patterns by Eclat rule mining algo.
Step 7. Eclat algorithm recommends the most common and most related movie to a user.
Step 8. Finally, calculate precision, recall, and F1-score by differencing average rating and similarity measures by total rating for all similarity measures.
Step 9. Exit

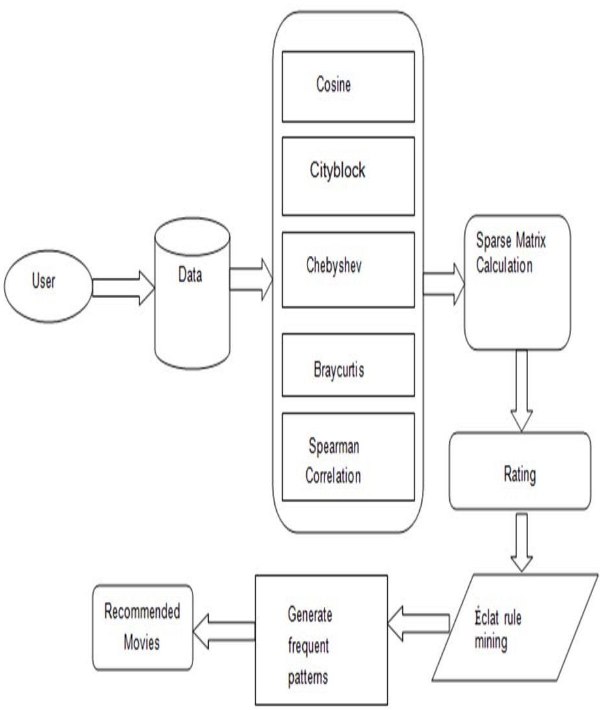

Figure 2: Proposed methodology block diagram.

IV. Result Analysis

This chapter gives an experimental analysis of our proposed system. Firstly, this section defines an experimental setup including a Dataset description. We performed a simulation using Python on Jupyter Notebook. Next, it presents the experiment results. The section involves an analysis of the effectiveness of the algorithm.

4.1 Dataset Description

We used free data for movies such as MovieLens, 1M dataset given in GroupLens Analysis, in our review. The dataset is public. It contains 100,000 movies on a scale of 1 to 5 for 943 users in 1682 movies. Dataset has been cleaned already. There is no need to preprocess datasets No pre-processing of the datasets is required. However, we have reformatted dataset files to suit our execution of a proposed algorithm. We took rating datasets from this available dataset (USerID, MovieID, and Rating, Timestamp), user datasets (UserID, Gender, and Age) as well as movie datasets from that dataset (MovieID, Title, Genres).

Table 4.1: Dataset Description.

Parameters	Value
Total no. of Rating	1,000,208
Total no. of Movie	3,706
Total no. of users	6,040
Total no. of Male users	4331
Total no. of Female users	1708

Figure 3 shows the precision, recall, & F1-score for all Cosine, Cityblock, Chevyshev, Braycurtis, and Spearman correlation measures similarity measures. Depending upon all observations, we discover that Spearman correlation outcomes in lesser precision rate (i.e. 0.8819), while cosine has highest precision rate (i.e. 0.887) in all 3 cases when compared to Cosine, Cityblock, Chevyshev, Braycurtis & Spearman correlation measures. Similarly, cosine results in the lesser F1-score rate (i.e. 0.903), while Cityblock and Chevyshev have the highest F1-score rate (i.e. 0.912) in all 3 cases when compared to Cosine, Cityblock, Chevyshev, Braycurtis & Spearman correlation measure

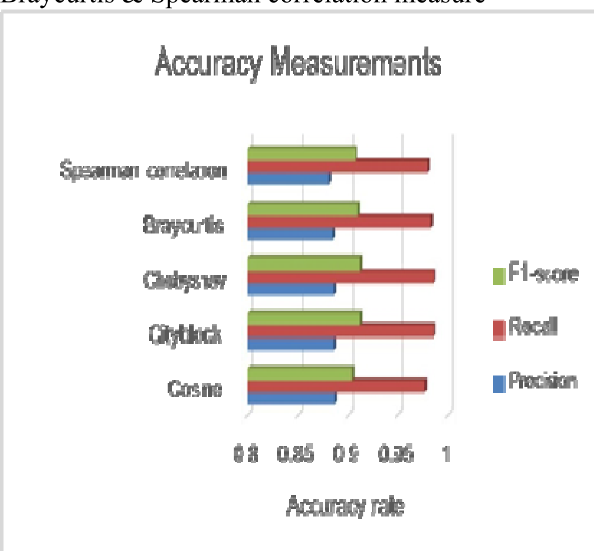

Figure 3: Recommendation Performance Comparison.

Figure 4 demonstrates that Item-based CF, RMSE rate is 3.61 (maximum) & MSE rate is 3.8 (highest) with Cosine similarity and RMSE rate is 3.34 (lowest) and MSE rate is 3.27 (lowest) with Braycurtis and Spearman correlation similarity. Depending upon all observations, we find that both Braycurtis and Spearman correlation similarity outcome in lesser error rate in all examples when related to Cityblock, Cosine, Chevyshev, Braycurtis & Spearman correlation measures.

Figure 4: Recommendation Error Graph.

V. Conclusion

We primarily based on movie recommendations in this proposed method using CF and RM methods. Reviews of this movie recommendation method are more perfect. First, an error is measured with five similarity measures to find the preview matrix and the result reveals that Chebyshev provides a lower error rate (i.e. 3.169). To mitigate new user issues and increase suggestion accuracy, Eclat rule mining is added to the prediction matrix. Finally, Eclat RM is calculated by recall, precision & F1 score measures. This was found that Cityblock and Chebyshev measure depended Eclat-RM has given improved F1 Score (i.e. 91.23% for both measures) and better precision (i.e. 88.66%) is given by cosine measures. The proposed MovieR-CERM has better recommendation accuracy also henceforth may fulfill users' needs & give them related movies they essential.

REFERENCES:

[1] D. Omtzigt, "Survey on Social inclusion :," vol. 5, no. January, pp. 1–34, 2009.

[2] C.-S. Wu, D. Garg, and U. Bhandary, "Movie Recommendation System Using

Collaborative Filtering," 2018, pp. 11–15, doi: 10.1109/ICSESS.2018.8663822.

[3] M. Liphoto, C. Du, and S. Ngwira, "A survey on recommender systems," Proc. - 2016 3rd Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2016, no. 09, pp. 276–280, 2017, doi: 10.1109/ICACCE.2016.8073761.

[4] N. Kalra, "Movie Recommender System using Collaborative Filtering," pp. 88–92.

[5] M. J. Omana, M. S. Monika, and M. B. Deepika, "Survey on Efficiency of Association Rule Mining Techniques 1," Int. J. Comput. Sci. Mob. Comput., vol. 64, no. 4, pp. 5–8, 2017.

[6] J. Dixit and A. Choubey, "A Survey of Various Association Rule Mining Approaches," 2014.

[7] B. Bhatt, P. J. Patel, H. Gaudani, and A. Professor, "A Review Paper on Machine Learning Based Recommendation System," Int. J. Eng. Dev. Res., vol. 2, no. 4, pp. 2321–9939, 2014.

[8] Y. G. Patel and V. P. Patel, "A Survey on Various Techniques of Recommendation System in Web Mining," Int. J. Eng. Dev. Res., 2015.

[9] S. G and P. G, "Survey Paper on Recommendation System using DataMining Techniques," Int. J. Res. Sci. Eng., vol. 2, no. 6, pp. 75–78.

[10] H. Ito, T. Yoshikawa, and T. Furuhashi, "A study on improvement of serendipity in item-based collaborative filtering using association rule," 2014, doi: 10.1109/FUZZ-IEEE.2014.6891655.

[11] A. S. Tewari, A. Kumar, and A. G. Barman, "Book recommendation system based on combine features of content based filtering, collaborative filtering and association rule mining," 2014, doi: 10.1109/IAdCC.2014.6779375.

[12] M. K. Kharita, A. Kumar, and P. Singh, "Item-Based Collaborative Filtering in Movie Recommendation in Real time," 2018, doi: 10.1109/ICSCCC.2018.8703362.

[13] L. Uyangoda, S. Ahangama, and T. Ranasinghe, "User profile feature-based approach to address the cold start problem in collaborative filtering for personalized movie recommendation," 2018, doi: 10.1109/ICDIM.2018.8847002.

[14] Z. Wang, T. Zhang, and H. Du, "A Collaborative Filtering Recommendation Algorithm Based on Density Peak Clustering," 2019, doi: 10.1109/CIS.2019.00018.

[15] F. Fessahaye et al., "T-RECSYS: A Novel Music Recommendation System Using Deep Learning," 2019, doi: 10.1109/ICCE.2019.8662028.

[16] M. Gupta, A. Thakkar, Aashish, V. Gupta, and D. P. S. Rathore, "Movie Recommender System Using Collaborative Filtering," 2020, doi: 10.1109/ICESC48915.2020.9155879.

[17] J. Guo, J. Deng, X. Ran, Y. Wang, and H. Jin, "An efficient and accurate recommendation strategy using degree classification criteria for item-based collaborative filtering," Expert Syst. Appl., 2021, doi: 10.1016/j.eswa.2020.113756.

[18] G. Suganeshwari and S. P. S. Ibrahim, "Rule-Based Effective Collaborative Recommendation Using Unfavorable Preference," IEEE Access, 2020, doi: 10.1109/ACCESS.2020.3008514.

[19] Z. Chen, X. Liu, and L. Shang, "Improved course recommendation algorithm based on collaborative filtering," 2020, doi: 10.1109/ICBDIE50010.2020.00115.