2019 4th International Conference on Information Systems and Computer Networks (ISCON)
GLA University, Mathura, UP, India. Nov 21-22, 2019

FRecS -Financial Recommender System

Parth Kapil
Bhagwan Parshuram Institute of Technology
New Delhi, India
parthkapil97@gmail.com,

Uday Shankar Acharya
Bhagwan Parshuram Institute of Technology
New Delhi, India
uday.sa97@gmail.com,

Yatika Bhardwaj
Bhagwan Parshuram Institute of Technology
New Delhi, India
yatikab17@gmail.com

Abstract— **People invest their income in different schemes and funds for future use and to fulfill their daily requirements. In today's fast-paced life, recommender systems are gaining a lot of attention because they can assist people in finding information about a product that they like. In the literature, no system till date is available for suggesting people how to save their money and also help in deciding if they are buying the right product. FrecS, the algorithm presented in this paper introduces a Collaborative Filtering (CF) approach, which is a technique used for generating high quality and accurate recommendations for the user. CF uses a subset of users who are called neighborhood users to get filtered recommendations for the current user. Moreover, this system utilizes the technique of simple heuristics to provide results, which in turn assist the user in getting a better recommendation without giving much details about them, which also helps them in securing their privacy. Online evaluation and the recommender technique is the basis of the empirical assessment in this system.**

Keywords— Collaborative filtering, Financial recommendation, Cosine similarity, Website recommendation, Shop recommendation

I. INTRODUCTION

With internet commerce growing at an exponential rate, it has become tiring and time consuming for consumers to find and buy things online [1]. With the system like online recommender, a lot of customized recommendations can be given to a consumer by gathering very little information about them. In a nutshell, the recommender system tends to aim for knowing the user's interest and then recommending them similar things based on their likings [7]. Companies like Amazon, Netflix, and Spotify have some great versions of recommender systems available nowadays [3]. The websites of these companies collect data from their users that are used afterward for recommending various songs, videos, and other products. The filtering techniques of the recommender system are categorized [15] into Collaborative[12], Content-based [11], and hybrid [13].

In this new approach, the 'social filtering' technique is used. Social filtering refers to filtering on the basis of how people in a similar social group feel about a particular product, what they like, and what they bought — for instance, both Patrick and Katie of the same age group like horror TV series. Patrick recently watched 'Conjuring' in amazon prime video and enjoyed it. Katie did not see the movie yet, but Amazon will recommend it to her because her interests are very much in common with Patrick. The social filtering method makes new predictions and interactions based on historical data from other similar users.

There are two different kind of approaches to obtain information in collaborative filtering: Item-based filtering [9,10] and user-based filtering [14].

Item-based filtering is an invention of Amazon. It finds and deduces how two things can be connected based on how many times they were bought together. For instance, if mechanical pencils and their respective lead packets appear in the cart multiple times, the next time when a different user tries to put a mechanical pencil in their cart, the recommender system will recommend similar items like lead packets according to thi algorithm.

In the user- based technique, the main filtering principle is different. It recommends products based on similar ratings and interests. This filtering method uses the distance between the similar preferences and likings of different users rather than the distance between the items. For instance, if Patrick and Katie like similar videos on facebook, the user-based technique will recommend Katie the videos that were liked by Patrick even though she has not seen them before.

The key implication mentioned above is that in both of the filtering systems, the system did not have prior information about the users or their interests. It only focuses on the number of times the items (mechanical pencil and lead packs) appeared together in the same cart and how both Patrick and Katie liked similar videos.

'Cognitive filtering' which is also referred to as content-based filtering. It is a technique in which various tags and attributes are assigned to the items. Therefore, the algorithm contains a database pool of all the content. Netflix exhibits the best example of this. Every movie on Netflix has a tag and genre associated with it. For instance, if Megan watches 'Moana' on Netflix, which has tags 'Children and family' and 'animated ' assigned to it. Next time she opens Netflix, similar Disney movies that have similar tags will be recommended to her based on the movie she has seen previously.

Collaborative filtering and content-based filtering both have limitations up to a certain limit. Therefore, a practical improvised approach is the hybrid model in which both approaches are fused together. The hybrid model is capable of providing more accurate and efficient results. The positive attributes of both techniques can be integrated into this model for a better recommending algorithm.

The sections are organized in this research paper as follows: In section II, the architecture of the financial recommender system (FrecS) is introduced, depicting various recommendation techniques that help in getting a list of multiple recommendations based on multiple filtering techniques. Then moving on to the analytic portion of the research, section III discusses the geometry of the dataset used and the pre-processing that was done to refine and segment the data so that it would work best in accordance with the algorithm. Section IV analyzes the quality and performance of the recommendations made by the recommender system. Finally, section V presents the conclusion, and Sec VI gives details about the prospects and future work.

II. FRecS Design and Architecture

The FRecS presents a collaborative filtering recommender system that recommends websites as well as shops to the user. Figure 2 depicts the whole architecture of the system(Frecs). The dataset (expenditure.csv), which is used for this research is first pre-processed(e.g., data cleaning and disambiguation). Whenever a new user requests a recommendation, a similarity array is computed between the current user and the existing users, which are present in the database using some similarity measures. The calculated similarity array is then rearranged in descending order. First, 'N' users in this rearranged array are the most similar users to the current user who has requested the recommendations. These first 'N' similar users are then stored in sim_user_set.

All the users in the sim_user_set are then filtered based on the positive ratings(i.e., rating>=3) that they gave to the shops and websites in the dataset. After filtering, these users are then finally stored in the final_recom_list, which is a list that contains users that are most similar to the current user who is requesting recommendations. At last, as a result, the current user is shown all the offline and online stores that are rated higher by similar users from the final_recom_list list.

Four properties from the dataset(expensiture.csv) were used for computing similarities between users. These properties were Location, Gender, Occupation, and Age. These properties are described as vectors of features. For computing, the similarity between users, various distance metrics can be used among these vectors. Some widely applied similarity measures are Euclidean[4], Pearson, cosine, etc. Cosine similarity is the similarity measure used here.

Cosine similarity is a metric that is used to find how similar the records are, regardless of their dimensions. Mathematically, it computes the cosine of the angle separating two vectors that exist in a multidimensional space [6,8]. The users are depicted as vectors, and the value of the cosine of the angle between the vectors is calculated with the help formula shown in equation 1. Vectors of two items with features are compared with cosine similarity function in figure 1.

$$u(c,s) = \cos(\overrightarrow{w_c}, \overrightarrow{w_s}) = \frac{\overrightarrow{w_c} \cdot \overrightarrow{w_s}}{||\overrightarrow{w_c}|| \times ||\overrightarrow{w_s}||} \quad (1)$$

[5]

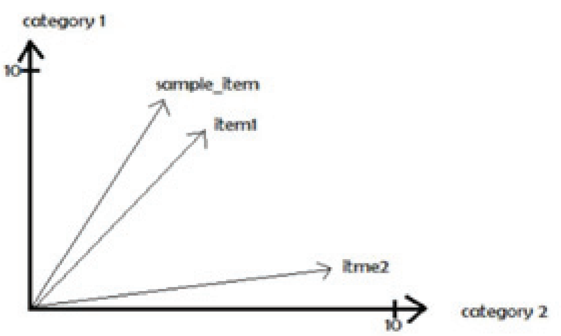

Fig. 1. Cosine similarity on coordinate plane

A. Locality Similarity:

The user's location (u_l) is given in the data provided by the user on the website. For finding the similarity between the current user and the other users in the system (v_{l_i} . . .. N) we use cosine similarity function.

$$l(u_l, v_{l_i}) = \cos(\overrightarrow{u_l}, \overrightarrow{v_{l_i}}) = \frac{\overrightarrow{u_l} \cdot \overrightarrow{v_{l_i}}}{||\overrightarrow{u_l}|| \times ||\overrightarrow{v_{l_i}}||}$$

[16]

B. Age Similarity:

The age of the users is divided into groups with a gap of 4 years. The age of the current user (u_a) is similar to the age of other users (v_{a_i}. . . . . . N) In the system if they belong to the same age group (A_i .. . . N).

$$A\ (u_a\ , A_i) = \{if\ u_a \in A_i,\ then\ 1\ else\ 0$$

C. Sex Similarity:

If the current user's sex (u_s) is same as the other user's

sex (v_{s_i} . . . . . . N) in the system then those users are considered to be similar.

$$S(u_s, v_{s_i}) = \{if\ u_s = v_{s_i}, then\ 1\ else\ 0\ \}$$

D. Occupation Similarity:

If the current user's occupation (u_o) is same as the

other user's occupation (v_{o_i} . . . . . . N) Present in the system then those users are considered to be similar.

$$O(u_o, v_{o_i}) = \{if\ u_o = v_{o_i}, then\ 1\ else\ 0\ \}$$

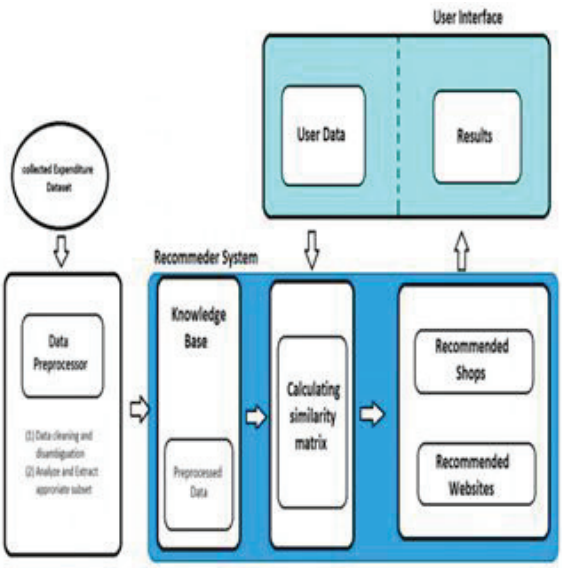

Fig. 2. FRecS system architecture

E. Prediction Generation:

The prediction score $P(u, v_i)$ of a user from the set of users (v_i... . . .. N), which shows how similar is that user to the current user (u) can be computed on the basis of a linear combination of similarity scores of the four attributes mentioned above:

$$P(u, v_i) = l(u_l, v_{l_i}) \times A(u_a, Ai) \times S(u_s, v_{s_i}) \times O(u_o, v_{o_i})$$

ALGORITHM

1. Start
2. Get the user data (u) containing Location, Age, Occupation, and Sex
3. Finding similarity of the current user (u) with the set of users present in the system (v_i. .. . .N) on the basis of four attributes which are locality, age, occupation, sex.
4. Multiplying the result of similarity scores of four attributes and storing it in the array for each user V_i present in the system.
5. Sorting the array in descending order.
6. Picking the first N users in the array. These are the first N similar users to the current user (u).
7. Classifying the top N similar user on the basis of the rating (i.e., rating > 3) given by other users as a recommendation set
8. Stop

III. DATA PROCESSING AND COLLECTION

We weren't able to find any dataset online, which would help us solve the problem set. Therefore, we conducted a survey from where we collected the dataset(Expenditure.csv). Data preprocessing techniques were applied to the dataset to eliminate duplicates and sparsity. The cleaned data was then used to seed the Frecs algorithm for generating similarity matrices.

A. Collection of user data

The data that was collected has 38 attributes. The data can be available from[18].

The proposed recommender system uses four major attributes out of total 38 attributes which comprise of:

- **Gender.** It refers to the gender of a user.
- **Age.** It refers to the age of the user.
- **Occupation.** It refers to the occupation of the user.
- **Location.** It referes to the vicinity where the user lives.

B. Data cleaning

Data cleaning is the first stepstone in processing the data. The extra unwanted data (null attribute entries) is of no help to the system. Getting rid of it saves time and efficiency in the data processing. The next step following the data cleaning is preprocessing, where a normal NLP[16] is applied in the result fields that are going to be exhibited to the user who has requested the recommendations.

C. Data Visualization

Some data exploratory data visualization is given bellow :

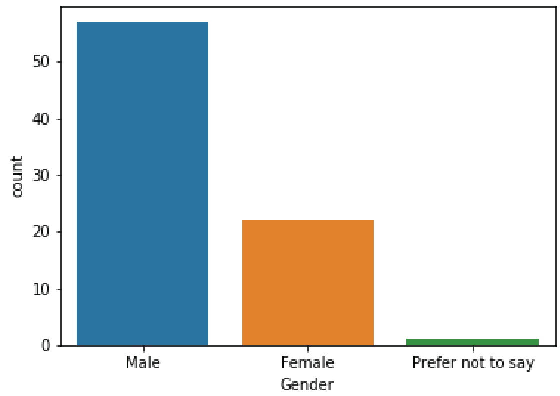

Fig. 3. The count plot for the user's gender in the database

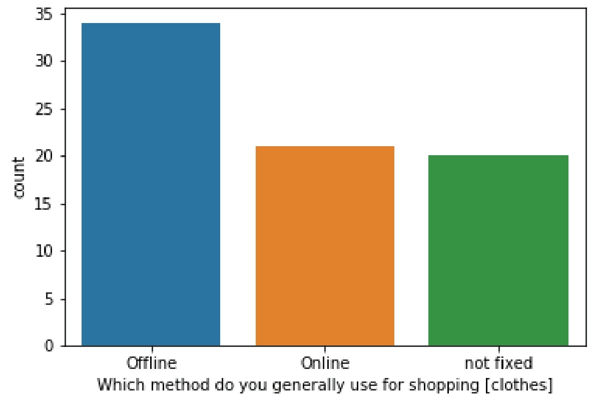

Fig. 4. The count plot for user's general method for shopping

Fig. 5. The Count plot for user's general method for shopping in hue of gender.

IV. EVALUATION AND ANALYSIS

B denotes the original preferences of the user, and the derived results are denoted as A. However, it is not possible to predict and calculate B accurately. The recommendations from the system will not always be what the user is actually looking for, and sometimes it can end up being some random recommendation. To resolve this, the principle of statistical estimation was opted, which uses the data of shops and websites that the user provided on the survey. With the help of it, the recommendations made by the system can be improved, and the degree of utility can be identified for a particular recommendation.

A. *Accuracy metrics:*

The actual value or the vicinity to the truth can be computed with accuracy metrics. It is measured as follows.

$$\text{Accuracy} = \frac{\text{Number of successful recommendations}}{\text{Number of recommendations}} \quad [17]$$

The accuracy achieved by Frecs is :

$$\text{Accuracy} = 68/83 = 0.819$$

$$\text{Accuracy (in percentage)} = 81.9\%$$

B. *Mean Absolute Error (MAE) metrics:*

Mean Absolute Error computes the mean absolute variation between each original preference B and the result X. Sum of accuracy and MAE is 1.

$$\text{MAE} = 1 - \text{Accuracy}$$

$$\text{MAE} = 1 - 0.81 = 0.19$$

V. CONCLUSION

This paper proposed a collaborative recommender algorithm which is able to recommend users websites and shops near there locality which provide great service and are highly rated by other similar users. To get these recommendations for a user, similarities are calculated between the current user and a set of users present in the system on the bases of four attributes and then similar users are marked. These marked users are then classified on the bases of ratings given by themselves. The final classified user's data is shown as a result to the current user who requested for the recommendations. FRecS ensures that the user gets a relevant recommendation at the same time we give them very good privacy which must be very important. Recommender systems involve an inherent trade-off between the accuracy of recommendations and the extent to which users are willing to release information about their preferences. Using online applications users may share or upload their personal information but this information is shared within the specific scope. The privacy of the information means exposure of the information within a bounded scope[5].

VI. FUTURE WORK

As it has been evidently seen in the past years that collaborative recommender systems were subjected to a lot of research and development but the amount of research done in the financial services are fairly less. However, the recent increase in the number of publications and new findings point to a prosperous future regarding advanced and upgraded financial recommender systems that could revolutionize many applications such as money investment and expenditure analysis.

ACKNOWLEDGEMENT

- We extend our sincere thanks to Assistant prof Charu Gupta and Assistant prof Amita Jain.
- We would also like to thank our project guide Assistant prof Mugdha Sharma.

REFERENCES

[1] H. Xu, R. Zhang, C. Lin, and W. Gan, "Construction of e-commerce recommendation system based on semantic annotation of ontology and user preference," Telkomnika Indonesian Journal of Electrical Engineering, vol. 12, no. 3, 2013.

[2] J. Lu, D. Wu, M. Mao, W. Wang and G. Zhang, "Recommender system application developments", Decis. Support Syst., vol. 74, pp. 12–32, 2015.

[3] G. Linden, B. Smith, and J. York, " Amazon.com Recommendations: Item-to-Item Collaborative Filtering, IEEE Internet Computing, vol. 7, no. 1, pp. 76 80, 2003.

[4] H. Tan and H. Ye, "A collaborative filtering recommendation algorithm based on item classification," in Proceedings of the 2009.

[5] Sanay Sharma, Aakriti Aharma, Yamini Sharma and Ms Manjot Bhatia, Recommender System Using Hybrid Approach, 2016.

[6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in Proceedings of the 10th International Conference on World Wide Web. Hong Kong, China: ACM Press, May 2001, pp. 285–295.

[7] Recommender system page on Wikipedia.[Online]. Available: http://en.wikipedia.org/wiki/Recommender_system

[8] A scalable collaborative filtering algorithm based on localized preference, vol. 1, 2008.

[9] Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor «Recommender Systems Handbook», 2011, 845p.

[10] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich, "Recommender Systems. An Introduction", 2011

[11] Iaquinta, M. D. Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino, "Introducing serendipity in a content-based recommender system," in International Conference on Hybrid Intelligent Systems, 2008, pp. 168–173.

[12] Hegde and S. K. Shetty, "Collaborative filtering recommender system," Esrsa Publications, vol. V4, no. 3, 2015.

[13] Salehi and I. N. Kmalabadi, "A hybrid attribute–based recommender system for e–learning material recommendation," Ieri Procedia, vol. 2, no. 4, pp. 565–570, 2012.

[14] Company Blog E-Content a "Quality Metrics ranking". Access: https://habrahabr.ru/company/econtenta/blog/303 458/

[15] Francesco Ricci and Lior Rokach and Bracha Shapira, Recommender Systems Handbook, Springer, 2011.

[16] Thomas Hornung, Cai-Nicolas Ziegler, Simon Franz, Evaluating Hybrid Music Recommender System, 2013

[17] Shruthi Bhat, K. Aishwarya, Item-Based Hybrid Recommender System for Newly Marketed Pharmaceutical Drugs

[18] https://drive.google.com/open?id=17SLasy_nomXCNBCf2mda_gO-bKPWx4_s