FINANCIAL RECOMMENDER SYSTEM

Charu Gupta, Palak Girdhar
Department of Computer Science and Engineering, Bhagwan Parshuram Institute of Technology, Delhi, India

Abstract-To meet their daily needs and for future use, people invest their money in variety of plans. In today’s era, recommender systems are getting lot of attention because it assist people in finding out the product and knowledge about the things they desire for. Recommender system provides ample number of similar options with various features. But no recommender system is available in the literature which recommends the product with the aim to save money. In this paper, we have designed an algorithm which will assist the people in smarter way. The algorithm is based on Collaborative Filtering (CF), which generates high quality and accurate recommendations to the users. In order to generate filtered suggestions for the current user, CF employs a subset of users it refers to as neighborhood users. Additionally, this system generates the outcome list using straightforward heuristics. This enables the user to receive better suggestions without disclosing a lot of personal information. The system's empirical evaluation is based on both the results of the online evaluation and the influence of recommender methods.

INTRODUCTION

The development of recommender systems technology has been fueled by the Web's growing significance as a platform for electronic and commercial transactions. The simplicity with which the Web enables users to provide information is a significant catalyst in this regard. People can provide the feedback about their likes and dislikes. For Example, considering a scenario of a content provider such as Hotstar. People can easily provide the feedback in terms of numerical values, showing the liking and disliking of the content, with just a single click of mouse. And this feedback can be advantageous to someone else.
Recommender system learns from the traces of user interactions and offers or recommends the personalized information to the user [2]. It aims to recommends the particulars items which are likely to be of interest of the user [7]. Amazon, Netflix, YouTube [3] are few examples of popular recommender systems. These platforms collect the data from the users and utilize the collected data for the recommendation. On the basis of gathered data, movies, playlists and videos are recommended to the user having same interest.
Recommender system suggest the items to user on the basis of their preferences. For this specific purpose, it uses various filtering techniques.

The filtering techniques of recommender system are categorized [15] into Collaborative, [12] Content-based [11] and hybrid [13].

Content based filtering suggests items of interest based on their distinguished features. For example, a person who is interested in reading news will be recommended next time with the similar words that are read before.
Collaborative filtering offers or recommends those particulars which are fetched on the basis of previous ratings of the similar interest’s users.
Hybrid techniques combines the features of both mentioned techniques to cope up the specific limitation of a single solution.

In the proposed approach, Collaborative Filtering is used. Recommendations for users are made on how similar other users liked the item. Collaborative filtering-based recommender systems uses rating for items provided by various users. It recommends the items to the target user who have not yet considered the suggested item but likely to have interest in the recommendation. The ratings are stored in a form of matrix, where rows specify the user’s ratings and columns denotes the item’s rating.

This filtering technique is sometimes also referred to as social filtering. Recommendation can be made to the other person having same interest or genre.

This method predicts new interactions on the basis of historical data.

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

Item based filtering technique was developed by Amazon originally. It computes the relationship between the two items that are bought together. For example, if bread and butter appear more often in the shopping bag or user history, it can be used for generating recommendations. Next time, when someone adds bread to his cart, system automatically recommend butter by learning the previous history.

The user-based technique differs from item-based filtering in that it calculates the distance between users based on their ratings or likes rather than on the goods themselves. Facebook's algorithms, for instance, can suggest videos that Angela liked to Sam even though Sam hasn't watched them

before if Sam liked certain videos on Facebook and Angela liked the related videos.

In both the techniques, system has no prior information about the relationship between the items and the users.

The system's sole concern is that either these products appeared in the same cart together or that people who have similar interests like it.

It also sometimes referred as cognitive filtering, content-based filtering is all about assigning tags and attributes to the items, so the algorithm of the system knows something about the content of each item in the database pool. Netflix is the most appropriate demonstration of this approach. Based on its genre, each Netflix movie is assigned a few tags. Brian, for instance, recently watched the movie "Bojack Horseman." It might be identified with the tags "animated," "adult," and "comedy." The next time, Brian will be given recommendations for programmes with similar tags to "Family Guy."

Both of the collaborative and content-based filtering have their shortcomings. Therefore, the hybrid approach combines multiple filtering techniques like collaborative and content-based to get the desirable and much more accurate results. One can achieve the best of both the worlds and get a much more accurate and precise data for the recommendations.

Further, the organization of the paper is follows: In section II, the architecture of the financial recommender system is introduced profiling various recommendation techniques that help in getting the result list of recommendations.

Section III discusses about the geometry of the dataset used and the pre-processing done to refine the data so that it would work best with the algorithm. Section IV analyzes the performance and the quality of the recommendations made by the recommender system (FRecS). Finally, section V presents the conclusion and Section VI elaborates the future aspects.

I. FRecS Architecture and design

The FRecS presents a collaborative filtering recommender system that recommends shops and websites to the user. The overall architecture of the system is shown in the following figure 2. The data set that has been used (Expenditure.csv) is first pre-processed (e.g. disambiguation and data cleaning). At runtime, whenever a new recommendation is requested by a user, a similarity array is calculated between this user and other users in the system using some similarity measures. This similarity array is then sorted in descending order. Afterward, top N users that are closest to the user who has requested the recommendations are marked in a sim_user_set.

Every user in the sim_user_set are classified with the help of the ratings, all the users who gave positive feedback (i.e. rating>=3) are then marked in the final_rec_list. Now with the available list, the top N similar users to the user that requested the recommendations are obtained. At last, the targeted user is shown all the shops and websites rated higher by the user in final_rec_list as the result.

For computing similarities, four attributes from the dataset were used which are Age, Occupation, Gender, and Location. These attributes are represented as vectors of features. The similarity between users can be calculated by using some distance metric between these vectors. Some commonly used similarity measures are cosine, Pearson, Euclidean [4] etc. The similarity measure used here is cosine.

Cosine similarity is a measure of calculating the similarity between two vectors of an inner product space that are non-zero that measures the cosine of the angle between them [6, 8]. The measure of the angle between two users, where users are represented as vectors gives out their cosine value [5] as shown in Equation 1. Vectors of two items with attributes are compared in cosine similarity function is shown in figure 1.

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

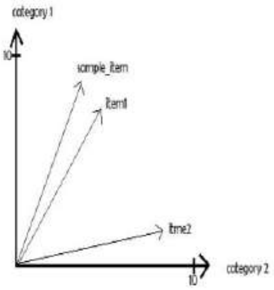


Figure 1: Cosine Similarity

1. 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 given in equation 2.

$$l(u_l, v_{l_i}) = cos\ cos\ (\overrightarrow{u_l}, \overrightarrow{v_{l_i}}) = \frac{\overrightarrow{u_l} \cdot \overrightarrow{v_{l_i}}}{\|\overrightarrow{u_l}\| \times \|\overrightarrow{v_{l_i}}\|} \quad [2]$$

2. 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 ($A_i \ldots\ldots N$).

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

1. 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.

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

2. 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.

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

3. Prediction Generation:

   The prediction score P (u,v_i) of a user from the set of users (v_i ....N). It shows how similar is that user to the current user (u) can be computed on the basis of linear combination of similarity scores of the four attributes mentioned above is given in equation 3:

$$P\ (u, v_i) = l\left(u_l, v_{l_i}\right) * A(u_a, A_i) * S\left(u_s, v_{s_i}\right) * O\left(u_o, v_{o_i}\right) \quad [3]$$

Following is the detailed algorithm: 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 rating (i.e. , rating > 3) given by other users as a recommendation set
8. Stop

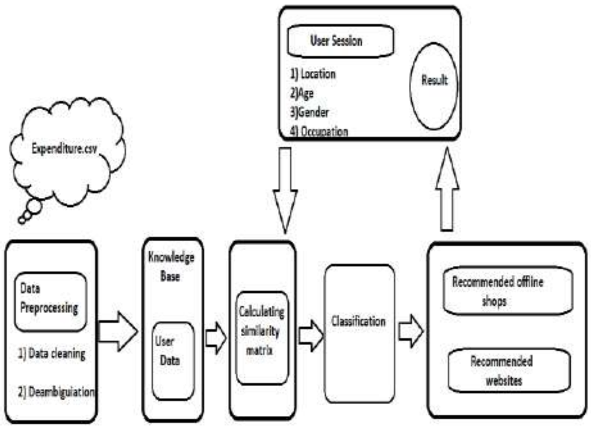


Figure 2: FRecS-FINANCIAL RECOMMENDER SYSTEM

II. Data collection and pre-processing

The data was collected via the google form because no such dataset was available online that could help in this particular case. The data was then cleaned for duplicates and for the sparsity. The clean data was then used to seed the two recommender techniques by allowing the computation of their similarity matrices.

1. *Collecting user data*

The data that is collected has 38 entries. You can get the data from [18]. In all of the 38 entries we used four entries for our recommender system which are:

The proposed recommender system uses four major attributes out of total 38. These are:

- **Occupation.** It shows the occupation of the user.
- **Age.** It shows the age of the user.
- **Gender.** It shows the gender of a user.
- **Location.** It shows the locality where the user lives.

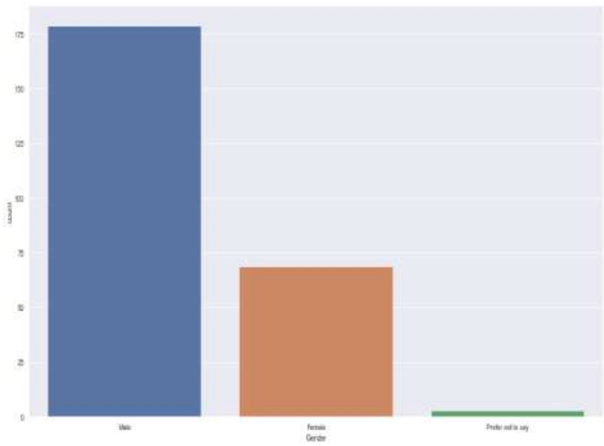
2. *Data cleaning*

The first step in preprocessing the data is the data cleaning, specifically removing data that is of use to the system (i.e. user having null entries in the attributes Age, Occupation, Gender, and Location). The next step in preprocessing is to apply normal NLP [16] in the entries that are to be shown as a result to another user requesting recommendations.

3. Data Visualization

Some data exploratory data visualization is given below:

- Figure 3 shows the count plot for the gender of the users in the database:



- Figure shows the count plot for the method user generally uses for shopping :

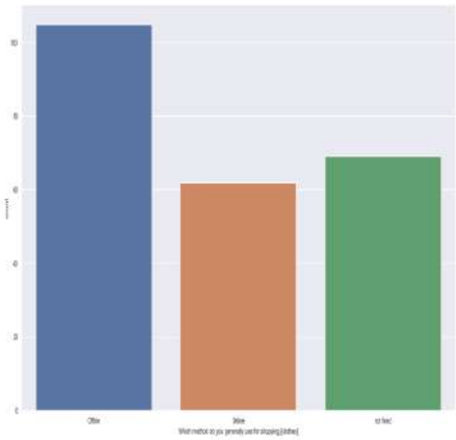


Figure 4: Shopping Trend

- Figure shows the count plot for the method user generally uses for shopping in hue of Gender:

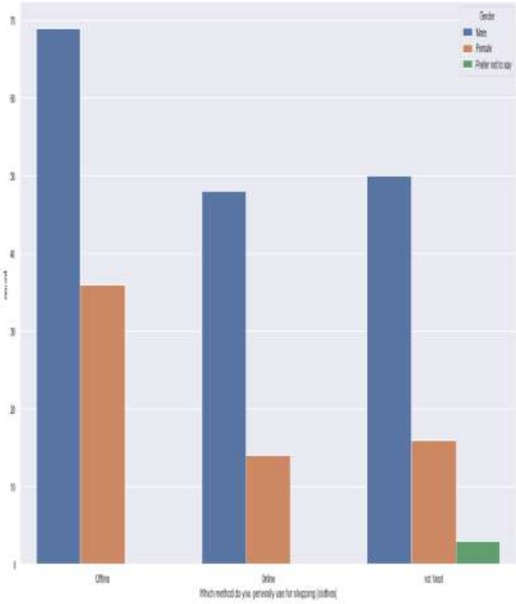


Figure 5: Shopping trend among males and females

III- Evaluation and Analysis

The results obtained by the recommender system will not always be in accordance with the user's preferences. The results obtained are denoted as A and user's original preferences are denoted as B with respect to the users. But it was not possible to accurately determine B. So, the principle of statistical estimation was opted which is based on the offline shops and websites that the user gave in the dataset. This way, the degree of utility of a particular recommendation to a user can be identified.
Accuracy metrics:

Accuracy metrics measure the quality of proximity to the truth or the true value achieved by the system. It is calculated from equation (4):

$$Accuracy = \frac{Number\ of\ successful\ recommendations}{Number\ of\ recommendations} \quad [4]$$

The accuracy shown by the FRecS is:
Accuracy = 68/83 = 0.819
Accuracy (in percentage) = 81.9%

Mean Absolute Error (MAE) metrics:
This metrics measure the average absolute deviation between each original preference B and the result X. the accuracy and MAE together add up to 1.

IV- 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 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. Our System ensures that the user gets a relevant recommendation on the same time we give them very good privacy which must be very important.

Recommender systems involve an inherent trade-off between 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

V- FUTURE WORK

It is observed that although a large amount of research and development has been done in the area of collaborative recommender systems, still a very small amount of research has been done related to the financial services. 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.

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] L. 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] A. Hegde and S. K. Shetty, “Collaborative filtering recommender system,” Esrsa Publications, vol. V4, no. 3, 2015.

[13] M. 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-Contenta. "Quality Metrics ranking".Access:https://habrahabr.ru/company/econtenta/blog/303458/

[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