International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Improvised Collaborative Filtering for Recommendation System

Shefali Gupta, Meenu Dave

Abstract: ***Collaborative filtering (CF) is one of the most important techniques of recommendation system and has been utilized by many e-commerce businesses to provide recommendation to its users. This paper sheds light on CF and its methods. This paper demonstrates a practical algorithm by leveraging data on user ratings for mobile phone devices and then provides recommendations to the target user based on the ratings given by similar users. It also elaborates an algorithm of CF that overcomes some of the common limitations faced by other algorithms. To explain the methodology of collaborative filtering this research paper looks at mobile phone data, especially the mapping of users (buyers) and the rating they provide for mobile phones they purchase. The model first evaluate multiple collaborative filtering techniques (variations of user based and item based filtering) by use of ROC curve and then provide recommendation to the user based on the best identified technique. Collaborative filtering is best utilized where the information on "users" and/or item is limited. For example, you can imagine the hotel booking website that provides recommendation to the website visitor, even though the user has never visited the website before (first time user). In such a situation as the information about user is limited the website algorithms are still able to utilize collaborative filtering methodology to provide recommendations.***

Keywords: Recommendation System, Collaborative filtering

I. INTRODUCTION

Recommendation systems are facilitators of decision making that leverages algorithm based on user preference and profile. These algorithms take advantage of similarity in the users or item and provides suggestions to the user regarding his/her next purchase (etc.). For example, next product to buy on Amazon, next movie to watch on Netflix or "people you may know" suggestion in Facebook.

Collaborative filtering has emerged as one of the key techniques of recommendation system. The basic crux of this technique is to provide recommendation to a user based on user – item mapping of a group of user.

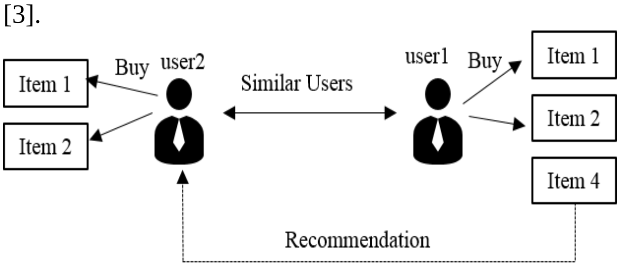
Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Shefali Gupta*, Ph.D. Scholar, Department of Information Technology, Jagannath University Jaipur, Rajasthan, India.
Prof. (Dr.) Meenu Dave, Ph.D., Department of Computer Science, Jagannath University Jaipur, Rajasthan, India.

For example, if your friends on Facebook like a particular music page, it is likely that Facebook with also recommend you the same page if you are interested in music as well [1].

Collaborative filtering has some advantages over content based filtering.

Firstly, it is content independent, i.e., it doesn't depend on item attributes or user profile for providing recommendation yielding error free results. Secondly, it takes into consideration real quality assessments as ratings are provided by users. Finally, It provides serendipitous recommendations as compared to content based filtering which recommends items similar to what user has already liked, it takes into different users interest resulting in diverse recommendations.

Collaborative filtering can broadly be classified into Memory based and Mod-el based Filtering techniques [2]. Memory based filtering takes into account en-tire dataset for providing recommendation to the user. It finds neighbors for an active user and generates a list of predictions of new items that can be suggested. Model based filtering uses training data to allow the system to identify complex patterns and make intelligent predictions for test data based on the learned model [3].



For example, if user1 has liked "item1", "item2" and "item4" and user2 who has liked "item1" and "item2" then the system will recommend "item4" as he/she is similar to user1.

II. RESEARCH METHODOLOGY

Figure gives general steps of collaborative filtering algorithm which need to be discussed in detail as follows:

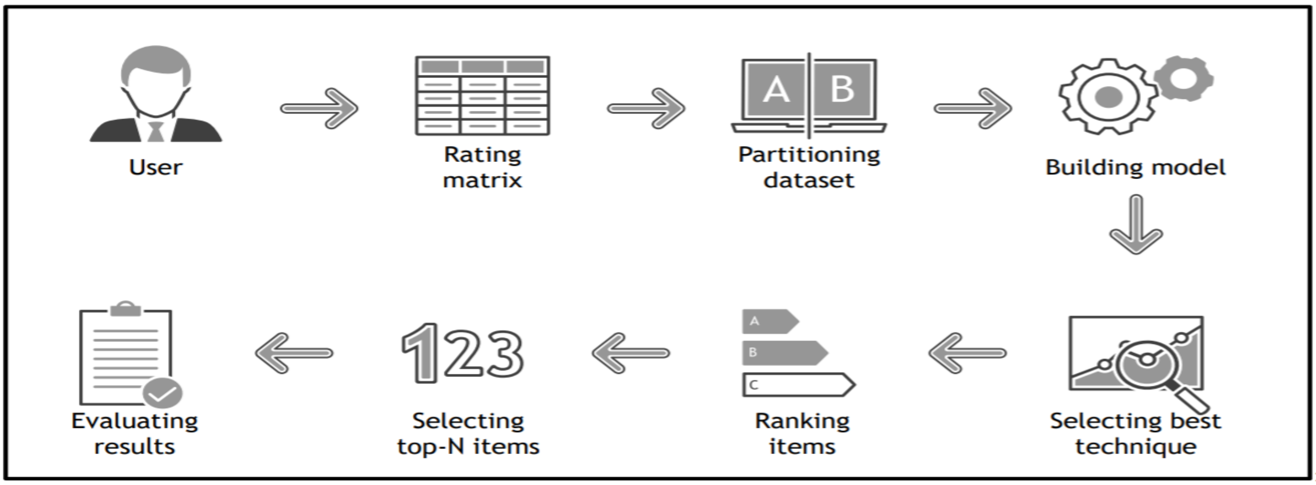
A. Data Collection

Data collection is the pre-requisite / fundamental of entire recommendation system. At this stage it is important to ensure the accuracy of the source of data. For the purpose of this algorithm mobile data was collected through a website and contains information that can be classified into 2 categories:

- Item attributes: This includes information regarding features of item like screen size, multi-touch yes or no, operating system, camera, memory, etc.

- User rating: This dataset includes rating given to a particular mobile phone by the user who has purchased it (or not). For example, a user may rate a phone with screen size 5” and RAM 2GB as 4(good) and phone with screen size 5.5” as 5(very good).

The rating is based on likert scale of 1 to 5, wherein 5 is the highest, 3 is median and 1 is the lowest rating an item can get.



In our collaborative filtering algorithm we have used user rating as a primary input for providing recommendations, i.e., item attributes are not referred to for this effort.

B. Data Preprocessing

Sourced data cannot be directly used on its face value as errors might creep up in data during the time of primary collection/recording or at any stage thereof. For example, some users might rate items arbitrarily (say rate all items as 5) to save time, or in the case of manual data collection, surveyor might incorrectly record rating shared by the user. There can also be technical glitches at the time of data collection that might add to the error in the dataset. These errors will lead to reduced reliability of the dataset and hence the accuracy of the results driven from the algorithm. Outlier detection methods [4] can limit data errors to some extent and must be used at this stage to ensure raw data is of optimal quality and can be processed by the algorithm.

Then, CF algorithms are executed to predict user preferences and recommend related items to the active user.

C. Algorithm for collaborative filtering

To create an effective algorithm for memory based collaborative filtering the user must ensure adequate focus on the following steps (as also highlighted in Figure 1 outlines).

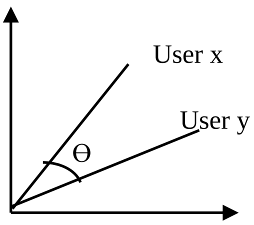
Step 1: Partitioning dataset

The user rating data collected (as discussed in point 3.1) and pre-processed (as discussed in point 3.2) can then be divided into two parts: Train dataset and test dataset [5]. Train dataset or learning dataset generally forms 80% of the data is used to build the model. It contains samples from the data and corresponding expected predictions. Test data which forms the remaining 20% of the data and is used to check the accuracy of the model built on train dataset. It contains only input samples from the data and we predict predictions for the same. There should be no overlap between the two datasets. Should there be any overlap, the accuracy of the model will be less. While 80:20 is a general split of this dataset, the user can select a different data mix based on the quality of data, number of observations and other such factors to yield optimal results.

Step 2: Building models

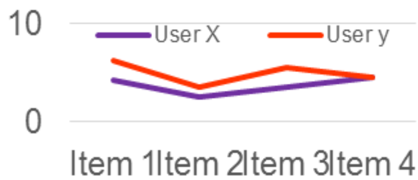
This step involves building multiple models based on different techniques of collaborative filtering. These techniques can broadly be classified under 2 heads – Item based collaborative filtering and user based collaborative filtering. Both of these techniques captures similarity between users or items by using various distance metrics:

- Cosine similarity index: It is used to measure similarity between two users by computing cosine angle between them [6].



$$\mathrm{sim}_{\mathrm{Cosine}}(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}$$

- Pearson correlation: It is used to measure linear correlation between two users.

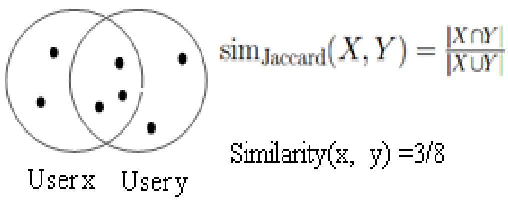


$$\mathrm{sim}_{\mathrm{Pearson}}(\mathbf{x},\mathbf{y}) = \frac{\sum_{i\in I} x_i y_i - I\bar{\mathbf{x}}\bar{\mathbf{y}}}{(I-1)s_x s_y}$$

Retrieval Number: F3513049620/2020©BEIESP
DOI: 10.35940/ijitee.F3513.049620
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

- Jaccard similarity index: It is measured as the size of the intersection divided by the size of the union of the samples.



Step 3: Selecting best technique

After building the model on each of the above technique, results are displayed using ROC Curve. A Receiver Operating Curve also known as ROC [7] sums up the interplay between the true positive rate and false positive rate for a predictive model using different probability thresholds. The AUC (Area under the curve) derived using ROC analysis is used to compare the performance of each model. The technique having maximum AUC will be picked for providing recommendations to the user.

User based and Item based similarity matrix [8] can also be calculated to find items similar to the item selected by the user and users similar to the active user. This will help in dealing with cold start problem [9, 12, 13] when we have limited data of users and items or if a user visits website for the first time.

Step 4: Ranking Items

Once the best model is picked, all the items are ranked according to their predict-ed ratings. Item with highest score will stand at the top.

Step 5: Selecting Top-N items

After ranking the items, User will be recommended Top-N [10] items where N is decided by the user before recommendation can be made.

Step 6: Evaluating results

The final step of the algorithm is to determine the accuracy of the result. For this, Precision and recall is used [11], wherein:

Precision is defined as:

Precision = true positive / (true positive + false positive)

It refers to percentage of results that are actually that are actually relevant (number of instances predicted as positive were actually positive). A high value of precision would mean less false positive values resulting in more accurate results.

Recall is defined as:

Recall = true positive / (true positive + true negative)

It refers to percentage of total actual results correctly classified by model. A high value of recall would mean less false negative values resulting in best results.

III. CONCLUSION

This paper provides an improvised algorithm for Collaborative filtering. It pro-vides recommendation to the users/buyers based on user ratings of the peer set. An algorithm has been proposed to explain the working of collaborative filtering in recommender systems. It first partitions the user ratings data in train and test dataset, builds a model on various collaborative filtering techniques, picks the model which is best and provides recommendation based on the best model choose.

Various distance metrics are also discussed in the paper that helps in finding users similar to the current user. Precision and recall are used to verify the accuracy of the model. In this time of big data, continuously improving collaborative algorithm will help in providing better recommendations to the user in less time and effort.

REFERENCES

1. H. Gao, Y. Xu, Y. Yin, W. Zhang, R. Li and X. Wang, "Context-aware QoS Prediction with Neural Collaborative Filtering for Internet-of-Things Services," in IEEE Internet of Things Journal, 2019.
2. M. Tang, Z. Zheng, G. Kang, J. Liu, Y. Yang, and T. Zhang, "Collaborative Web service quality prediction via exploiting matrix factorization and network map," IEEE Trans. Netw. Service Manage. vol. 13, no. 1, pp. 126-137, Mar. 2016.
3. S. Agrawal and P. Jain, "An improved approach for movie recommendation system," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, 2017, pp. 336-342.
4. Z. Yang, B. Wu, K. Zheng, X. Wang and L. Lei, "A Survey of Collaborative Filtering-Based Recommender Systems for Mobile Internet Applications," in IEEE Access, vol. 4, pp. 3273-3287, 2016.
5. Bandyopadhyay S., Thakur S.S, "Product Prediction and Recommendation in E-Commerce Using Collaborative Filtering and Artificial Neural Networks: A Hybrid Approach," In: Mandal J., Sinha D. (eds) Intelligent Computing Paradigm: Recent Trends. Studies in Computational Intelligence, vol 784. Springer, Singapore,2020
6. I. Bartolini, Z. Zhang, and D. Papadias, "Collaborative Filttering with personalized skylines," IEEE Trans. Knowl. Data Eng., vol. 23, no. 2, pp. 190-203, Feb. 2011.
7. A. Gogna and A. Majumdar, "A comprehensive recommender system model: Improving accuracy for both warm and cold start users," IEEE Access, vol. 3, pp. 2803-2813, Dec. 2015.
8. B. Bulut, B. Kaya, R. Alhajj and M. Kaya, "A Paper Recommendation System Based on User's Research Interests," 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, 2018, pp. 911-915.
9. Y.Mo, J.Chen, X.Xie, C.Luo, andL.T.Yang, "Cloud-basedmobilemultimedia recommendation system with user behavior information," IEEE Syst. J., vol. 8, no. 1, pp. 184-193, Mar. 2014.
10. Pereira N., Varma S.L. (2019) Financial Planning Recommendation System Using Content-Based Collaborative and Demographic Filtering. In: Panigrahi B., Trivedi M., Mishra K., Tiwari S., Singh P. (eds) Smart Innovations in Communication and Computational Sciences. Advances in Intelligent Systems and Computing, vol 669. Springer, Singapore
11. Azzopardi J. (2018) Item-Based Vs User-Based Collaborative Recommendation Predictions. In: Szyma?ski J., Velegrakis Y. (eds) Semantic Keyword-Based Search on Structured Data Sources. IKC 2017. Lecture Notes in Computer Science, vol 10546. Springer, Cham
12. Lika, Blerina & Kolomvatsos, Kostas & Hadjiefthymiades, Stathes. (2014). Facing the cold start problem in recommender systems. Expert Systems with Applications: An International Journal. 41. 2065-2073. 10.1016/j.eswa.2013.09.005.
13. L. Yanxiang, G. Deke, C. Fei and C. Honghui, "User-based Clustering with Top-N Recommendation on Cold-Start Problem," 2013 Third International Conference on Intelligent System Design and Engineering Applications, Hong Kong, 2013, pp. 1585-1589

AUTHORS PROFILE

Shefali Gupta, is a Ph.D. Scholar in Information Technology at Jagannath University Jaipur, Rajasthan. She received her Master's degree in Computer Science (MSc. Comp Sc...) From University of Delhi in 2016. She has two paper publications in national and international conferences, and one in an international Journal.

His research interests are Data Mining, Machine Learning and Artificial Intelligence.

Prof. (Dr.) Meenu Dave, M.Tech., Ph.D. (Computer Science) has taught Computer Science in different capacities at a number of Engineering Colleges and Institutes. She ha extensive experience in teaching Cloud Computing, Artificial Intelligence, data science,Knowledge Management and Data Mining at the post graduate level. She has also authored several research papers in the specified areas which have been published in leading journals.