

Group Recommendation for Cold Start Users Using Hybrid Recommendation Technique

Harleen Kaur, Gourav Bathla

Abstract: Recommender system is an data retrieval system that gives customers the recommendations for the items that a customer may be willing to have. It helps in making the search easy by sorting the huge amount of data. We have progressed from the era of paucity to the new era of plethora due to which there is lot of development in the recommender system. In today's scenario the interaction between the groups of friends, family or colleagues has increased due to the advancement in mobile devices and the social media. So, group recommendation has become a necessity in all kinds of domains. In this paper a system has been proposed using the group recommendation system based on hybrid based filtering method to overcome the cold start user issue which arises when a new user signs in and he/she doesn't have any past records. So, the recommender system does not have enough information related to the user to recommend an item which will be of his/her interest. The dataset has been taken from the MovieLens is used in the experiment.

Keywords: Cold start problem, Group recommendation system, Hybrid filtering approach.

I. INTRODUCTION

In this era of data overload and abundance, people find it hard to select a particular item for themselves from the large dataset they see on media this kind of issue is known as information overload. So, to overcome this problem recommender system came into play.

The principle behind the recommender system is to suggest all kinds of information about the elements to the users according to their actions[1]. The recommender system focuses on the personalized approvals from the users according to their preferences whether they are explicitly done or implicitly. Explicitly refers to the ratings given by the customer to a particular item as per his or her likings. Implicitly refers to the browsing or the purchases made by the user[2]. The three main approaches of recommender systems are content based, collaborative filtering based and the hybrid based filtering approaches. The other name for content based filtering approach is cognitive filtering. In the content based filtering approach it recommends items on the footings of the user's description and the user's item's description. The most widely used collaborative filtering technique uses the opinion such as ratings of the other person which are quite similar to the user who has to be recommended. In case hybrid recommender system it combines both the other filtering approaches, content based and collaborating based filtering approaches[4]. But there are some limitations in recommender system such as cold start, low scalability, data sparsity and the shilling attacks. Cold start complication emerges when a new user signs in and there are no past

records of the user[1]. To get rid of this problem we used group recommendation system using hybrid filtering approach because it will amplify the accuracy and the efficiency of the recommendation. Content based filtering method faces lower accuracy while recommending to the cold start users and on the other hand collaborative filtering method go wrong while recommending to the cold start users even though it frequently suggests accurate recommendations otherwise. Hybrid recommendation tries to aggregate this information and get the accurate results. In today's scenario the interaction between the groups of friends, family or colleagues has increase to the advancement in mobile devices and the social media. So, group recommendation has become a necessity in all kinds of domains. In case of group recommender system the factors that influence the system are social relation, background, trust, relation, similar interests etcetra[3].

A. Types of group

The groups are mainly classified into two based on their composition are as follows:

- 1) Homogeneous Groups

A homogeneous group comprises of the members that have similar interests among themselves.

- 2) Heterogeneous Groups

A heterogeneous group consists of the users that have diverse interest among their group members. It is quite hard to perform group recommendation on the heterogeneous groups due to the fact that all members have different liking and to satisfy all of them on a particular item is very challenging.

Groups can be further classified on the basis of the interaction between the group members. They are as follows:

- 3) Established Group

The group members of the established group who have come together explicitly because of their similar interest they share.

- 4) Occasional Group

In occasional group the members of the group are heterogeneous but occasionally they will have a common interest. For example all the members are planning for a vacation at some particular place therefore they will have a common aim.

- 5) Random Group

People of random group have shared an environment for a particular time without the linkage among them.

Revised Manuscript Received on July 22, 2019.

Harleen Kaur, Chandigarh university, Gahruan, Mohali, Punjab, India
Gourav Bathla, Chandigarh university, Gahruan, Mohali, Punjab, India

For example in the case of public transport the passengers meet unlike their interest at a particular time[3].

B. Preference Acquisition

The quality of suggestions in the recommendation system have high dependence on the data that has been used is as preferences is of high quality or not. In group recommender system the data acquisition can be classified as follows:

1) Explicit data Acquisition

In explicit data acquisition the customers are provided with the survey that they have to go through and then for a particular item their preferences are stored on the basis of the rating scale. The collection of the preferences can have the individual item preferences as well as for the group of items too. The drawback of this acquisition is the data sparsity because all the group members will not rate every item available.

2) Implicit data acquisition

In implicit data acquisition the users donot give their preferences directly. The data that has to be collected about them is done indirectly by tracking the user behavior. For example how much time does the user spent on a particular website, gender, age, occupation, relationships among the group members.

3) Negative feedback data acquisition

Along with the positive preferences negative preferences are equally important. In case of group recommendations the negative feedback is very important because it helps in avoiding some particular items being disliked by some of the group members. The influence of the users with the minority opinion is enhanced by the negative preferences system[3].

II. RELATED WORK

Group recommendation has become the profound concern of researchers and is now a days an effective research area. The prime focus of this paper[1] is to handle the cold start user problem by using the group recommender system. The groups are detected automatically by two different methods proposed, one is the analogous feature wise and other is the social performance wise. Social performance wise group detection method helps in solving the cold start user problem. This paper[2] provides the guidance over the group recommendation using the collaborative techniques along with some group modeling strategies. They used k-means algorithm for cluster formation. This paper[3] surveyed the state-of-the-art in group recommendation systems used in different domains. They also gave an overview on the aggregation models and the customer preference models.

In this paper[5] the author have automatically detected groups by using clustering algorithm in the group recommendation system. They used two different datasets to produce their desirable results. Different modeling strategies and algorithms have been used and compared at the end. In[6] recent research work is focused on addressing the cold start complication by giving suggestions to the new users who don't have any past history related to their preferences. They recommended on the basis of top-N recommendations. In the research work by[7] a method is proposed to resolve the data

sparsity and the cold start issue with the help of collaborative filtering techniques. They have incorporated the users likings on the non functional features as the additional information which will lead to the increased accuracy in recommendation system. Naïve Bayes classifier have been used on to the cold start problem by[8]. They have combined content based and collaborative based filtering methods to propose a method to suggest elements to the new person. In paper[9] a similarity measure is proposed to suggest thongs to group of people. The author have used collaborative filtering approach. This paper[10] provides an efficient solution to get rid of cold start user problem. They used the hidden interest of the group members to recommend to the new or the target user the one does not have any past history regarding the preferences. The prime focus of this paper[11] is to review all the existing techniques that have been proposed using collaborative filtering method to overcome the cold start drawback. This paper[12] gives an insight on the challenges that are being faced in the group recommender system while recommending to the group members.

III. PROPOSED FRAMEWORK

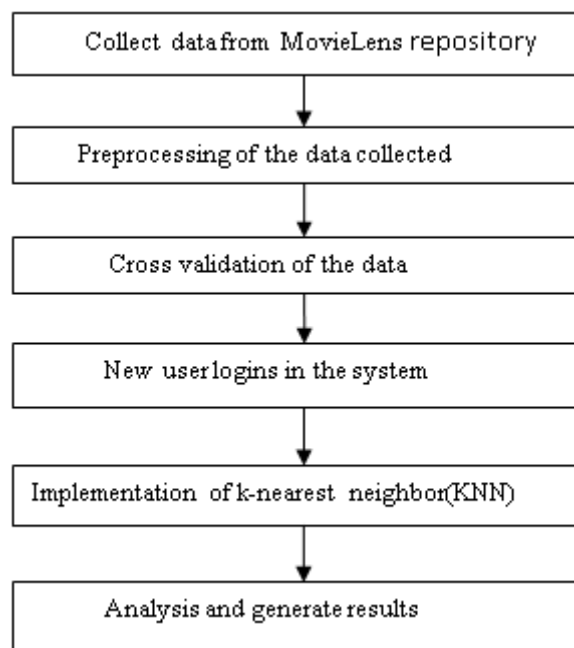


Fig 1: Process flow diagram

Step 1: Collection of data: Firstly, we search the dataset in MovieLens repository. The dataset comprises of 1,000,209 ratings for 3,900 movies given by 6,040 MovieLens customers.

Step 2: Pre-processing of usable data from collected data: The main purpose is to capture the useful and non-trivial data from the unstructured text data. In other words, it is considered as a renovation of raw material (Information) in a comprehensible format. The raw data have many variables which add the noise to distort the accuracy[14].

Table 1: Comparison of different research works done

| Sr.no | Author | Filtering method | Algorithm and Similarity measure | Cold start problem | Sparsity problem | Evaluation metrics |
|-------|--------------------|------------------|----------------------------------|--------------------|------------------|-----------------------------------|
| 1 | Ghodsad[1] | C.F | - | ✓ | × | - |
| 2 | Katarya[2] | C.F | K-means | × | × | Recall, Precision, RMSE, MAE |
| 3 | Boratto[5] | C.F | Clustering | × | ✓ | RMSE |
| 4 | Yanxiang[6] | Hybrid | Clustering, Vector cosine method | ✓ | × | Precision, RMSE, MAE |
| 5 | Fletcher[7] | C.F | Pearson Correlation | ✓ | ✓ | MNAE |
| 6 | Schein[8] | C.F, C.B | Naïve Bayes | ✓ | × | ROC curve |
| 7 | Ortega[9] | C.F | - | × | × | Precision, Recall |
| 8 | Hawashin[10] | C.B | - | ✓ | × | Searching time, space consumption |
| 9 | Proposed Technique | Hybrid | KNN, Euclidian Distance | ✓ | × | Precision |

dataset into two pieces, so that the model can be trained and tested on different data. It uses the data effectively.

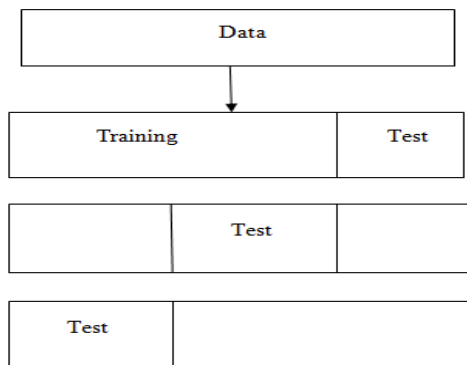


Fig 2: Cross validation

Step 4: New user logs in the system: The new customer logs in to the system by entering his/her features like age , gender, occupation and zipcode. On the basis of these features he/she will be assigned a group and further the items will be recommended on the behalf of the likings of the group members.

Step 5: Implementation of k-nearest neighbor(KNN) Classifier: Nearest neighbor classifier is the instance based classifier, which is quite popular among the common approaches to collaborative filtering. It finds out the number of k nearest points against the training records. KNN classifier allocates the class label to the closest class label. In implementing the KNN classifier the challenge is to opt for the value of k. The value of k varies, if its small the classifier will be prone to the noisy points and if the value is large, many points will be added from the neighborhood. Even though the KNN classifier is known to simple and intuitive but it shows excellent accuracy and is susceptible to the changes[13]. In this paper clusters are formed on the basis of some of the user like age, gender, occupation and the zipcode. If the euclidean distance is less between two users that means they will fall into same group.

A. Similarity measure

Euclidean distance

The most commonly used and the simplest distance measure is the Euclidian distance:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (1)$$

Where n is the number of attributes and x_k and y_k are the kth attributes of the data objects x and y respectively. Less the distance between the target user and the other users are more closely they are related.

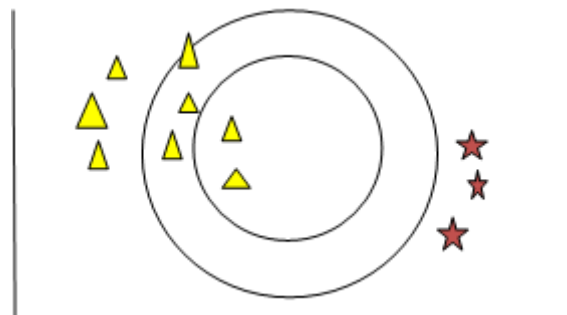


Fig 3: KNN

Triangles and the stars represent different class labels and the question mark is the new example to classify.

Step 5: Analysis and generate results: While implementing K-nearest neighbor for forming the groups. The new user is assigned a group according to the user and movies will be recommended with the rating above four. Precision and recall have been used as the evaluation metrics, the values obtained are precision is and recall is

IV. EXPERIMENT ANALYSIS

A. Software used

Python has been used in the implementation. It is the most practiced programming language as it is a open source software.

It was created by Guido Van Rossum in 1991. It works on different platforms such as Windows, Mac, Linux, etc. It has a very simple syntax familiar to the English language. It makes use of whitespace indentation, instead of curly brackets or keywords, to limit the blocks. After some statements the indentation is expanded and contracted when the current block is to end. Python 3.7 has been used while implementing.

B. Dataset Description

The dataset is taken from MovieLens archive. The dataset comprises of 1,000,209 ratings for 3,900 movies given by 6,040 MovieLens customers. For ratings a 5-star scale has been used. User data contains age, gender, occupation and zipcode. Gender stands for "M" for male and "F" for female. Age is elected from a certain range for example 1 for Under 18, 18 for 18-24, and so on. Occupation is to be choosed from the choices such as 0 for not specified, 1 for academic/educator, 2 for artist, and so on. Movie information contains the name of the movie, its id, and the genre. Genres added are Action, Adventure, Animation, Children's, Comedy, Crime etcetra.

Table 2: MovieLens dataset statistics

| | | |
|--------------------|-----------|---------|
| Dataset statistics | | |
| Users | 6,040 | Movies |
| 3,900 | | Ratings |
| | 1,000,209 | |

C. Evaluation Metrics

Precision

Precision is defined as the proportion of items that the user admires, to the total number of items that are recommended.

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \quad (2)$$

We have compared our approach with the existing approach. In this paper[6] they addressed the cold start user issue by giving suggestions to the new users, as they don't have any past history related to their preferences. They recommended on the basis of top-N recommendations. The value of precision they got was 0.651. We have improved the value of precision to 0.782 by implementing our proposed methodology.

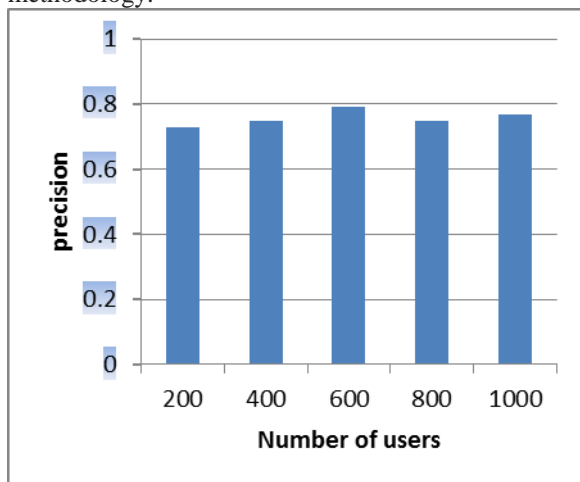


Fig 4: Precision for MovieLens dataset

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a hybrid recommendation approach for cold start users, these users don't have any past information regarding the preference for any item. We have used user based clustering for the formation of the groups and then give them the top recommendations. The new user get assigned to a particular group and is recommended on the basis of the likes and dislikes of the group members. The future work will further improved by using different distance measures and using the large scale data.

REFERENCES

- Ghodsad, P. R., & Chatur, P. N. (2018, August). Handling User Cold-Start Problem for Group Recommender System Using Social Behaviour Wise Group Detection Method. In 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE) (pp. 1-5). IEEE.
- Katarya, R., & Verma, N. (2017, December). Automatically detection and recommendation in collaborative groups. In 2017 International Conference on Intelligent Sustainable Systems (ICISS) (pp. 218-222). IEEE.
- Dara, S., Chowdary, C. R., & Kumar, C. (2019). A survey on group recommender systems. Journal of Intelligent Information Systems, 1-25.
- Patel, A., Thakkar, A., Bhatt, N., & Prajapati, P. (2019). Survey and Evolution Study Focusing Comparative Analysis and Future Research Direction in the Field of Recommendation System Specific to Collaborative Filtering Approach. In Information and Communication Technology for Intelligent Systems (pp. 155-163). Springer, Singapore.
- Boratto, L., & Carta, S. (2015). ART: group recommendation approaches for automatically detected groups. International Journal of Machine Learning and Cybernetics, 6(6), 953-980.
- Yanxiang, L., Deke, G., Fei, C., & Honghui, C. (2013, January). User-based clustering with top-n recommendation on cold-start problem. In 2013 Third International Conference on Intelligent System Design and Engineering Applications(pp. 1585-1589). IEEE.
- Fletcher, K. K. (2017, June). A Method for Dealing with Data Sparsity and Cold-Start Limitations in Service Recommendation Using Personalized Preferences. In 2017 IEEE International Conference on Cognitive Computing (ICCC) (pp. 72-79). IEEE.
- Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002, August). Methods and metrics for cold-start recommendations. In Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 253-260). ACM.
- Ortega, F., Hurtado, R., Bobadilla, J., & Bojorque, R. (2018). Recommendation to groups of users using the singularities concept. IEEE Access, 6, 39745-39761.
- Hawashin, B., Mansour, A., Kanan, T., & Fotouhi, F. (2018, October). An efficient cold start solution based on group interests for recommender systems. In Proceedings of the First International Conference on Data Science, E-learning and Information Systems (p. 26). ACM.
- Revathy, V. R., & Anitha, S. P. (2019). Cold Start Problem in Social Recommender Systems: State-of-the-Art Review. In Advances in Computer Communication and Computational Sciences (pp. 105-115). Springer, Singapore.
- Delic, A., & Masthoff, J. (2018, July). Group Recommender Systems. In Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (pp. 377-378). ACM.
- Amatriain, X., Jaimés, A., Oliver, N., & Pujol, J. M. (2011). Data mining methods for recommender systems. In Recommender systems handbook (pp. 39-71). Springer, Boston, MA.

14. Yang, W., Fan, S., & Wang, H. (2018, October). An item-diversity-based collaborative filtering algorithm to improve the accuracy of recommender system. In 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (pp. 106-110). IEEE.
15. Yang, X., Guo, Y., Liu, Y., & Steck, H. (2014). A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41, 1-10.
16. Nehete, S. P., & Devane, S. R. (2018, August). Recommendation Systems: past, present and future. In 2018 Eleventh International Conference on Contemporary Computing (IC3) (pp. 1-7). IEEE.
17. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web* (pp. 291-324). Springer, Berlin, Heidelberg.
18. Ambulgekar, H. P., Pathak, M. K., & Kokare, M. B. (2019). A Survey on Collaborative Filtering: Tasks, Approaches and Applications. In *Proceedings of International Ethical Hacking Conference 2018* (pp. 289-300). Springer, Singapore.
19. Guo, L., Liang, J., Zhu, Y., Luo, Y., Sun, L., & Zheng, X. (2018). Collaborative filtering recommendation based on trust and emotion. *Journal of Intelligent Information Systems*, 1-23.
20. Thorat, P. B., Goudar, R. M., & Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications*, 110(4), 31-36.
21. Mustafa, N., Ibrahim, A. O., Ahmed, A., & Abdullah, A. (2017, January). Collaborative filtering: Techniques and applications. In 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE) (pp. 1-6). IEEE.
22. Han, D., Li, J., Li, W., Liu, R., & Chen, H. (2019). An app usage recommender system: improving prediction accuracy for both warm and cold start users. *Multimedia Systems*, 1-14.
23. Zhu, Y., Lin, J., He, S., Wang, B., Guan, Z., Liu, H., & Cai, D. (2019). Addressing the Item Cold-start Problem by Attribute-driven Active Learning. *IEEE Transactions on Knowledge and Data Engineering*.
24. Belém, F. M., Heringer, A. G., Almeida, J. M., & Gonçalves, M. A. (2019). Exploiting syntactic and neighbourhood attributes to address cold start in tag recommendation. *Information Processing & Management*, 56(3), 771-790.
25. Silva, N., Carvalho, D., Pereira, A. C., Mourão, F., & Rocha, L. (2019). The Pure Cold-Start Problem: A deep study about how to conquer first-time users in recommendations domains. *Information Systems*, 80, 1-12.
26. Li, J., Zhang, K., Yang, X., Wei, P., Wang, J., Mitra, K., & Ranjan, R. (2019). Category Preferred Canopy-K-means based Collaborative Filtering algorithm. *Future Generation Computer Systems*, 93, 1046-1054.

AUTHORS PROFILE

Harleen Kaur, is pursuing M.E in computer science in Chandigarh university, Punjab. She has complemented her B.Tech in information technology from Gurur gobind singh indraprastha university, New Delhi. She is a research scholar and have published 1 paper in a reputed journal. Her area of interest are data minin, Machine learning and data analytics. She can be contacted at harleenkaur3663@gmail.com.

Gaurav Bathla, is working as Assistant Professor at Chandigarh University, Punjab, India. He has 12 years of teaching experince. He has completed M.E from Delhi College of Engineering. He is GATE qualified with All India Rank 59. He is an active researcher and published 10 research papers in reputed journals and 5 research papers in International conferences. His areas of interest are Big Data, Machine Learning, Deep Learning, Data analytics and Programming Langauges. He may be contacted at gouravbathla@gmail.com.