International Journal of Advanced Research in Engineering Technology & Science ISSN: 2394-281

Email: editor@ijarets.org Volume-4, Issue-1 January- 2017 www.ijarets.org

SOLVING COLD-START PROBLEM IN RECOMMENDER SYSTEM WITH SEGMENTS OF HUMAN POPULATIONS

Gaurav Agarwal
Research Scholar
Uttarakhand Technical University
Dehradun

Dr. Himanshu Bahuguna
Professor
Shivalik College of Engineering
Dehradun

Dr. Ajay Agarwal
Professor
Krishna Institute of Engineering & Technology
Ghaziabad

ABSTRACT: Recommender systems have been used immensely commercially, scholastically and economically, recommendations created by these systems intend to offer relevant useful items to users. Several ways have been recommended for providing users with recommendations utilizing their rating history; most of these approaches suffer from new user problem (cold-start) which is the initial lack of items ratings. This paper suggest propose new user segment information to give suggestions as opposed to utilizing rating history to stay away cold-start problem. We present a framework for evaluating the usage of different segment qualities, such as age, gender, and occupation, for recommendation generation. Experiments are executed using Movie Lens dataset to evaluate the performance of the proposed framework.

I. INTRODUCTION

In the recent years, recommender system has been utilized colossally commercially, scholastically and economically for providing users with recommendation about products, services, or information which match their inclinations and interests. These suggestions are recommended by the system to guide user in a personalized way based on user's historical preferences to discover unseen items among an awesome gathering of items stored on the system. Recommender systems are utilized in various spaces to customize its applications by recommending items, such as books, movies, melodies, eateries, news articles, jokes, among others.

Researchers have proposed several approaches for building recommender systems which offer items distinctively to users based in view of a particular supposition keeping in mind the end goal to coordinate their interests. By and by, all proposal approaches have qualities and shortcomings that ought to be considered while picking the most reasonable way to deal to implement. In this manner, hybrid recommenders are commonly utilized for joining at least two suggestion approaches together acquiring better execution and fewer drawbacks [1].

The recommendation systems types can be distinguished into two most used approaches:

Collaborative Filtering: Collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). It refers that users with comparative tastes will rate items likewise. It endeavors to discover users having comparable rating history to the target user (user who requires recommendations), building an area (neighborhood) from which the recommended items are generated.

Content-based Filtering: This approach tries to recommend items that are similar to those that a user liked in the past (or is examining in the present). However, the above methodologies had been tended to suffer from new user problem, refers as cold-start problem, which is having initial lack of ratings when a new user join the system [3]. Since both methodologies supposition are based upon user's ratings history, this issue can significantly influence adversely the recommender performance because of the inability of the system to deliver significant suggestions [4]. Hence, an option sort of input (alternative) is required to be acquired explicitly from users to be utilized for suggesting recommendations instead of ratings.

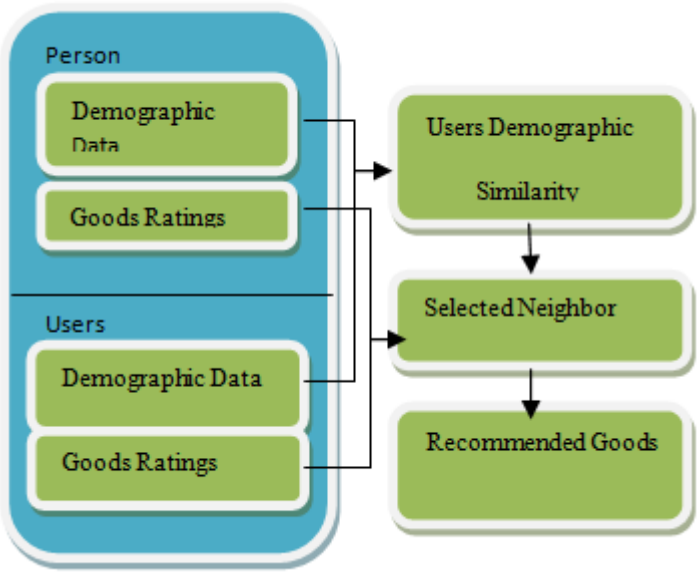

Fig. 1 - Demographic- based approach.

Another approach had been introduced which utilizes user segment information as an alternative input for recommendation in recommender system which is refer as segment-based approach. User Segment approach, as shown in **Fig.-1**, proposes using users' segment data stored on their profiles (i.e. Gender, age, location … etc.), it expect that users with comparative segment qualities will rate goods comparably i.e. similarly. This approach gets gathering of user having comparably segment quality(s) shaping an area from which newly suggested i.e. recommended items are generated.

II. WORK UTILIZING STATISTIC CHARACTERIZATION BASED APPROACH IN RECOMMENDER SYSTEM

Researchers apply the hybridization of statistic and collaborative approaches in the recommendation system. Here, k-nearest neighborhood approach had been applied which figures the similitude scores between the target i.e. objective client and other clients shaping an area i.e. neighborhood, increasing the points or scores of users having comparable ratings and statistic characterization qualities (each segment characteristic had been assessed along comparable ratings separately) [2,6]. While another research work shown another altered version of k-nearest neighborhood by adding a user segment vector to the user profile, the similarity estimation consider both ratings and user segment vector [7].

Hence, the recommended framework uses the statistic data to resolve the new user "cold-start" problem. The framework goes for assessing the impact of statistic qualities on the user rating, to help the recommender framework architect to enhance proposals i.e. recommendation quality for new users. The framework had been examined using a movie dataset to evaluate the generated recommendations accuracy and precision [2].

III. STATISTIC CHARACTERIZATION EVALUATION FRAMEWORK FOR RECOMMENDER SYSTEM

The statistic characteristic-based approach performs three phases: Input (data) similarity i.e. comparability calculation and recommendation calculation (as shown in **Fig. 2**). Such as:

(a) Data input phase: this phase holds the statistic characteristic data of new target user's (user requires recommendations) and also ratings and statistic characteristic data of the rest of users.

Similarity calculation phase: This phase uses statistic characteristic data of the users to acquire various users having comparable user segment data to **the target i.e. objective user shaping an area.**

(b) Recommendation calculation phase: This phase acquires items which have been ordinarily positive-evaluated by neighborhood users to be proposed to the target i.e. objective client.

For example, the below **Table I** shows the user segment data of four users; every user has four segment qualities i.e. qualities (age, nation, gender and occupation). Let us assume that David is a new user who request suggestion or recommendation, the system needs to ascertain the likeness amongst David and different users in light of the chose traits. The similar i.e. likeness computation yield relies on upon the way the system translates how users are comparative, if users having a similar occupation are comparative then Samvid is similar to David**,** else if users having the same nationality and gender are comparative then Herain and Kamil are similar to David. In this manner, the selection of qualities influences the comparability computation output which subsequently affects the result of recommendation.

Table I - Example of Users Segment Data

Name	Gender	Occupation	Country	Age
David	M	Businessman	India	15
Herain	M	Doctor	Germany	54
Samvid	F	Businessman	India	36
Kamil	M	Teacher	France	28

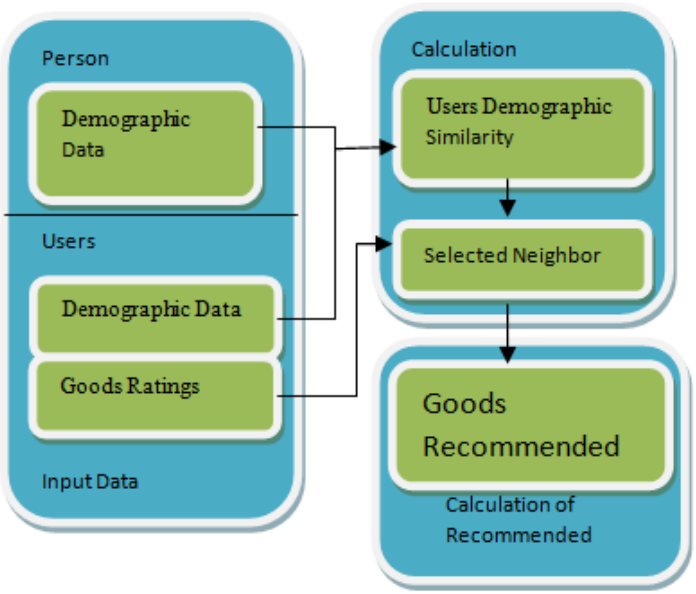

Fig. 2. Statistic characteristic based approach for new users.

The suggested framework comprises of many modules: data source, quality analysis, splitting dataset and recommendation generation. Data source has all data information about users stored. Quality examination module meets expectations around examining the real statistic characteristic situations, the dissemination from claiming values over those dataset and legitimacy of utilizing these qualities for recommendations. Part dataset module parts developing and testing for each substantial quality by evacuating all ratings of some users from preparing documents and ratings. Furthermore recommendation suggestion system extracts most frequent goods appeared in the preparing document (rated eventually clients hosting comparative quality of the concealed users) recommending them of the new clients (hidden users), the testing document will be utilized for assessing those accuracy of the recommendations contrasted with the hidden ratings [2].

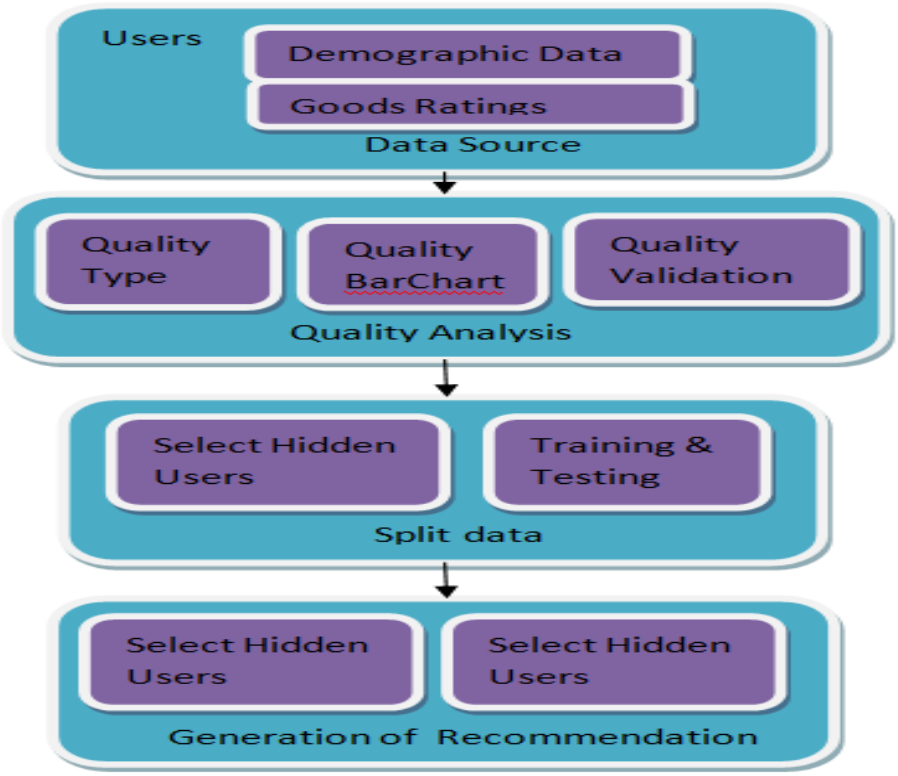

Fig.3 - Statistic characteristic evaluation framework.

IV. Experimental Methodology

The framework experimentally tried different things utilizing the publicly accessible information about Group Lens motion picture recommender system. This dataset required been utilized toward large number researchers; a few analysts utilized those dataset on execute their encounter experience [2, 7] same time others utilized those Movie Lens dataset should investigation the state-of-art for recommender frameworks applying collective approach diverse systems [11]. Additionally, Group Lens gives Different variants of the dataset, For example, Movie Lens. 100k, Movie lens 1M datasets.

Table II: Movielens Dataset Information

Movielens Dataset Files	File Quality Description
User	The user file include demographic information for 811 persons.“user id/gender/age/pincode”
Goods	The goods file include information for movies “movie id/ title/release data/URL/ Adventure/action/drama/comedy/Animation/crime/ fantasy/docum-Entary/horror/Thriller/musical/R-Omance.”
Data	The data file include 1000000 ratings By 811 users on 1561 goods. “User id/goods id/rating/stamp”

A. Information Source

The dataset utilized is MovieLens 1M; it comprises from claiming 1000000 ratings which were estimated by 811 clients looking into 1561 movies. Every user required rated in any event 100 movies; the ratings would allocated numerically from 1(bad) should 6(very good). Table II demonstrates about MovieLens dataset files utilized within the experiment.

B. *Quality Analysis*

The quality investigation module determines the kind and worth ranges of the user segment qualities for Movielens dataset, demonstrated previously in Table III. That point those frequency of each quality esteem is computed indicating those amount of users hosting comparative value; fig.4 illustrates those histogram about Movielens user segment qualities but for zip code which required just 148 low frequent duplicated values. Afterwards, that module validates those qualities that might a chance to be utilized for recommendations by checking those following conditions:

1) Invalid Value Range: It happens at the point when some ranges of the user segment quality need low frequency, for example, the recurrence of age quality "less for equivalent twelve year old" has only five users in this range Fig. 4 (b), therefore, those framework won't have the capacity should show proposals for new users who fall in this range.

2) Invalid Quality Value: It exists at worth from claiming a trait need a ambiguous meaning, for example, such that occupation quality values none and other in Fig. 4 (c)), In more than one client required their occupation filled similarly as none alternately other it does not mean that they are hosting comparative taste alternately will rate things.

3) Invalid user Quality: User qualities may be acknowledged invalid at its values would exceptionally sparse, for example, such that pin code value the greater part about its values are unique same time couple new low frequencies.

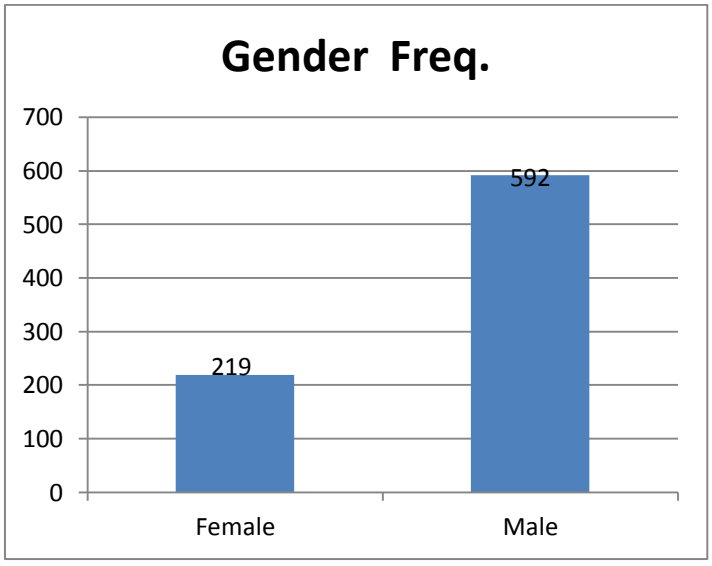

Fig. 4(a)

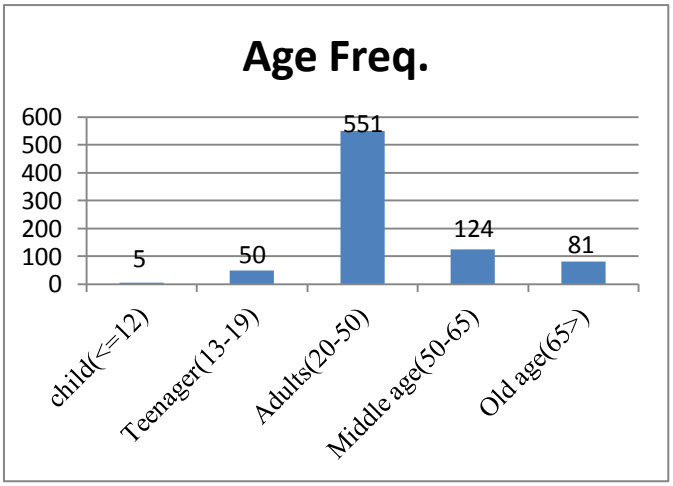

Fig. 4(b)

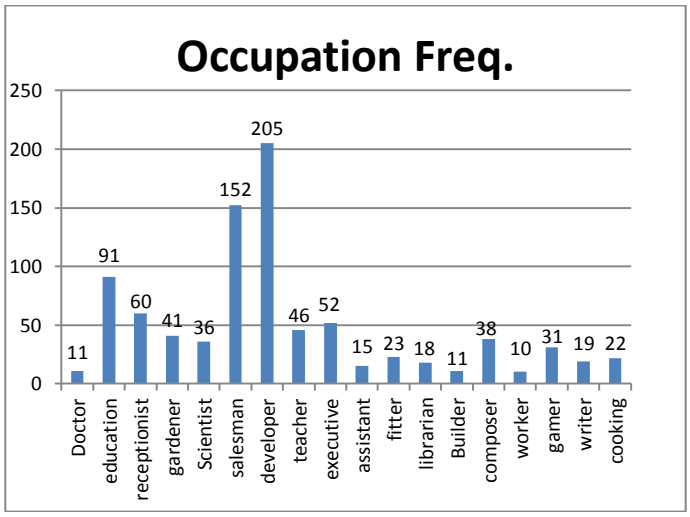

Fig. 4(c). Demographic Quality

C. Splitting Dataset

Table III: Segment Qualities Types

Attribute Name	**Data Type**	**Value Ranges**
Gender	Character	M, F
Age	Number	8-80
Occupation	Text	20 Occupations
Pincode	Text	800 distinct value

Split dataset module makes preparing and testing files to every of the three substantial user qualities (age, gender, occupation) furthermore their substantial values (excluding "age" invalid reach and "occupation" two invalid values) with a chance to be assessed. Dataset requires selecting amount for clients will conceal their appraisals including them to trying dataset. Table IV illustrates those amount of users whom their appraisals will a chance to be evacuated from preparing dataset (40 users for every segment qualities); our analysis just the vast majority four incessant values of occupation qualities: Student, Programmer, Home maker, Educator, will a chance to be acknowledged same time whatever remains of occupations will be excluded for purpose of diminishing those number of trials.

D. Recommendation Generation

The suggestion era module uses those preparing record from claiming each quality to calculate the frequency for know things rated by users hosting comparative quality worth. To instance, those module employments gender orientation preparation document on figure the frequency of goods rated eventually by female and the other way around to male orientation.

V. Conclusion and Future Work

In this we have introduced a best structure to assessing statistic characteristics accessible to recommender frameworks datasets on be utilized for recommending important things on new users. Those frame might have been analyzed utilizing MovieLens dataset, those experimental test effects of the dataset demonstrated that constantly on qualities need very nearly the same impact. Conclusively, it appears that the statistic characteristic information in the MovieLens dataset doesn't impact distinctively once users thinking.

Further examine could a chance to be performed on upgrade those results, for example, making more than person preparing and testing dataset make assessed what's more assemble the normal of the comes about. Also a higher level of pen recommendation can be obtained by relating the pen genres to statistic characteristic.

REFERENCES

1. R. Burke, "Hybrid web recommender systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, ed., Springer, 2007, pp. 377-408.
2. Laila Safoury and Akram Salah, "Exploiting User Demographic Attributes for Solving Cold Start Problem in Recommender System," Lecture Notes on Software Engineering, Vol. 1, No. 3, pp. August 2013.
3. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, pp. 734-749, June 2005.
4. J. B. Schafer, D. Frankowski, and J. Herlocker, "Collaborative Filtering Recommender Systems," in *The Adaptive Web: Methods and Strategies of Web Personalization*, P. Brusilovsky, A. Kobsa, and W. Nejdl, ed., Springer, 2007, pp. 291-324.
5. D. Almazro, G. Shahatah, L. Albdulkarim, M. Kherees, R. Martinez, and W. Nzoukou. "A Survey Paper on Recommender Systems," *arXiv preprint arXiv: 1006.5278*, Dec. 2010.
6. A. Said, T. Plumbaum, W. E. De Luca, and S. Albayrak, "A comparison of how demographic data affects recommendation," in *Proc. 19th international conference on User modeling, adaption, and personalization*, 2011.
7. M. Vozalis and K. G. Margaritis, "Collaborative filtering enhanced by demographic correlation," in *Proc. IAI Symposium on Professional Practice in AI, of the 18th World Computer Congress*, 2004.
8. M. Jones, "Introduction to approaches and algorithms for Recommendation systems" December 12, 2013 [online] available at: https://www.ibm.com/developerworks/library/os-recommender.
9. G. Shani and A. Gunawardana, "Evaluating recommendation systems," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, ed., Springer, 2011, pp. 257-297.
10. Wikipedia, "Recommender System" [online] available at: https://en.wikipedia.org/wiki/Recommender_system Hybrid_Recommender_Systems.
11. M. Jones, "Introduction to approaches and algorithms for Recommendation systems" December 12, 2013 [online] available: https://www.ibm.com/developerworks/library/os-recommender.
12. Xiao-Lin Zheng, Senior, Chao-Chao Chen, Jui-Long Hung, Wu He, Fu-Xing Hong, and Zhen Lin, "A Hybrid Trust-Based Recommender System for Online Communities of Practice", IEEE TRANSACTIONS on Learning Technologies, Vol. 8, No. 4, October-December 2015.
13. Yuan Cheng, Jaehong Park, and Ravi Sandhu, "An Access Control Model for Online Social Networks Using User-to-User Relationships", IEEE TRANSACTIONS on Dependable and Secure Computing, Vol. 13, No. 4, July/August 2016, Pp. 424-425.
14. S. Walters, L. Cooper, K. Jubas, S. Butterwick,"Hard/soft formal/informal work/learning: Tenuous/persistent binaries in the knowledge-based society", *J. Workplace Learn.*, vol. 20, no. 7/8, pp. 514-525, 2008.
15. E. C. Wenger, W. M. Snyder, "Communities of practice: The organizational frontier", *Harvard Bus. Rev.*, vol. 78, no. 1, pp. 139-146, 2000.
https://en.wikipedia.org/wiki/Recommendersystem
16. L. Candillier, K. Jack, F. Fessant, and F. Meyer. "State-of-the-art recommender systems," in *Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling*, M. Chevalier, C. Julien, and C. Soule-Dupuy, ed., IGI Global, 2009, pp. 1-22.