

A study to the cold start problem based on click stream data in the recommender system.

¹ ANURAG SINGH, ² Dr. SUBHADRA SHAW

¹ Research Scholar, ² Asst. Professor,
Department of Computer Science and Application,
AKS University, Satna(M.P.), India.

Abstract : Recommender systems have to deal with cold start problems as new users and new items are always present, where systems are unable to recommend relevant items to the users due to the unavailability of adequate information about them. In literature, many researchers have addressed this problem by collecting missing information, but their approaches differ in the way they collect missing information. This paper tries to propose a concept to solve the cold start problem by combining collaborative filtering, demographic approach, and click stream visit count.

Keywords: Recommendation system, clickstream, collaborative filtering, demographic analysis.

I. INTRODUCTION:

The growing importance of the website as an electronic and business medium has served as a driving force in the development of recommendation technology. Recommendation systems (RS) are software tools that recommend items of interest to users based on their preferences. RS's are important features of the e-commerce system [1]. Researchers have suggested a few ways to create a provider system that offers things differently to users based on certain assumptions to match their interests. However, all recommendation methods have strengths and weaknesses. Therefore, hybrid recommendations are often used to combine two or more recommendation methods that carry better performance and fewer issues [2]. The recommendation system types can be divided into two most commonly used recommendation approaches: The First is collaborative filtering (CF) technique recommends the items to the target user based on past preferences of other users with similar tastes [3]. CF helps users make decisions that support the opinions of others. The second is a content-based recommendation system, in which case, descriptive attributes of the items are used to make recommendations. The term "content" refers to these descriptive. With content-based methods, users' ratings and buying behavior are combined with content information in the items [4]. The main problem is the CF recommender methods are "data sparsity" and the "cold start" problem, without full knowledge, it is difficult for the CF model to recommend correctly [5]. Three types of cold start problems can be identified by (I) new users' recommendations (II) recommendations of new items and [III] recommendations on new items for new users [6].

1.1 Click Stream:

Click stream indicates the user's path through a web site. Blog files store and maintain all click stream knowledge. This knowledge are often terribly useful in providing the effective recommendation. Nice quality recommendation systems won't simply facilitate in fulfilling client's inclinations for associate item however additionally in enhancing feels and attracting new buyers [18]. Researchers trying to overcome the above-discussed problem. Cold start is the most important challenge for RS's. In this paper, a new hybrid approach is proposed based on the combination of CF and demographic approach with clickstream count, to solve the cold start problem of the new item type. we focused on solving the item side as well as user side cold start problem.

2. LITERATURE REVIEW:

In this section, we review some of the works related to our proposed approach.

D.Almazro et al.[7] Introduced a hybrid recommendation system using a demographic-based and CF approach to the movie domain using demographic data to reinforce the advice suggestion process. **Qing Li et al.**[8] In their explain how the clustering techniques can be applied to item-based CF framework try to solve the cold start problem. **A.Said et al.**[9] This paper introduced the transformation to the k-nearest neighbor, which calculates the similarity scores between the target user and other users which creates the neighborhood and increases the user's scores with similar ratings and demographic characteristics. **D.Zhang et al.**[10] proposed recommendations for starting the cold using a collaborative filtering system, which includes composition, filtering, and forecasting steps. They assume that users are followed by sensors such

that each user has their own location, which is currently considered an object. The user matrix of the item has been compiled and compiled to identify users who have a significant impact on the recommendation. Predictive measures are performed by taking a hybrid between item-based and user-based filtering methods. **Braunhofer et al.[11]** have taken a little bit different problem and so tried to predict a new user. They developed context –Aware recommender using user demographic details. **H. Sobhanam & et al.[12]**, In this paper, authors have explained the fusion recommender system that how association rules could be used to expand user profiles so that a new user problem is addressed and the clustering technique could group items based on similarity to address the new item problem. **Jindal & et al.[13]** Authors explain a hybrid method comprised on the basis of the grouping of collaborative and content-based filtering is proposed which does its work in two modes: offline mode, online mode. Offline mode made clusters out of given dataset using K-means clustering algorithm. While online mode uses the rule obtained from decision tree from offline mode and identifies the cluster user falls in and uses average ratings by users in that cluster. **N.K Mishra and et al.[14]**, In their research paper have used market basket analysis to address the cold-start problem. The Market basket Analysis to address the cold start problems. It's used for association rule mining on new users. **Xuan Nhat Lam and et al.[15]**, Authors worked on solving cold start problem has proposed a hybrid model which is grounded on the inspection of two probabilistic aspect models that utilizes users information, including age, gender, and job. Using these factors, a triadic aspect model was formed centred on the clue that people having identical features will also share similarities in interests. This model is claimed to resolve the user-side cold start problem to a certain extent.

3. PROPOSED WORK:

This work is based on information theory. In this paper, a new hybrid approach is proposed based on the collaborative filtering, demographic similarity between users in order to solve cold start problem of new user type. Also a new approach is proposed based on the click stream data in order to solve the cold start problem of new item type. The proposed work can be divided into some parts. Fig-1 shows the approaches of recommendation system.

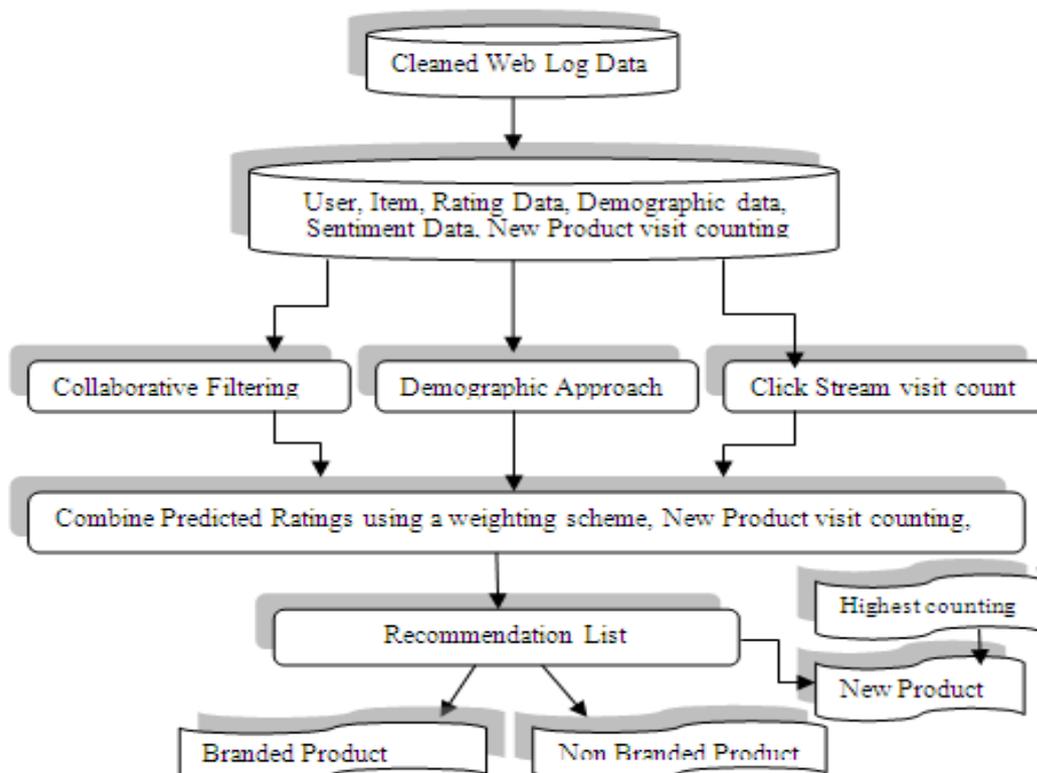


Fig.-1 Approach of Recommendation system

Step-I: The unstructured data will be converted into the structured form after applying pre-processing and a user-item matrix will be generated.

Step-II: In this step we generate user-product rating Matrix, consider a set of n users $U = \{u_1, u_2, u_3, \dots, u_n\}$ and a set of m items $I = \{p_1, p_2, p_3, \dots, p_m\}$. Let R_{ij} ,

$1 \leq i \leq n, 1 \leq j \leq m$, denotes the rating value of user I on product j . If $R_{ij} \neq 0$ indicates that the user I have not rated the product j [16]. The rating matrix R is given in the following.

$$R = \begin{matrix} & \text{I1} & \text{I2} & \dots & \text{Im} \\ \text{U1} & \text{R11} & \text{R12} & \dots & \text{R1m} \\ \text{U2} & \text{R21} & \text{R22} & \dots & \text{R2m} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \text{Un} & \text{Rn1} & \text{Rn2} & \dots & \text{Rnm} \end{matrix} \quad [16]$$

Step-III: The main concept of collaborative filtering is to calculate similarity products or users. At present, scholars have proposed many similarity measures, cosine similarity, Adjusted Cosine Similarity, Pearson coefficient, and other methods. Among them, the Adjusted Cosine Similarity is the best among many calculation methods. In this approach, we analyze the Adjusted Cosine Similarity method.

$$sim(i, j) = \frac{\sum_{c \in I_{ij}} (R_{ic} - \bar{R}_i)(R_{jc} - \bar{R}_j)}{\sqrt{\sum_{c \in I_{item}} (R_{ic} - \bar{R}_i)^2} \sqrt{\sum_{c \in I_{item}} (R_{jc} - \bar{R}_j)^2}} \quad [17]$$

Step-IV: In this approach, the demographic profile of customers is considered in computing item ratings. This consideration is based on the fact that demographics contribute to differences in people's tastes or preferences. A clustering algorithm such as K-means partitions users based on user demographic Data using Euclidean distance as the distance measure. The prediction for an item using demographics based user clusters will be taken as the average rating of the user cluster for that item.

Step-V: This step used a clickstream scheme for a new item. There is not much information about the new item. So such types of items have not any customer ranking. How to the recommendation of that new item. For that first new item upload with all features and company name of a particular e-commerce website. Whenever registered or unregistered users visit on the website, then count the click on that product of every user. Along with e-commerce websites also count clickstream data of that product on social media. Thereafter follow the CSDC (clickstream data count) Approach. Fig-2 Shows the flowchart of Click stream visit count.

Algorithms are:

Input: click stream data of new Item.

- I) CSD=0 ,
- II) IF (CSD >=1) then goto step IV else goto click stream count
- III) If (user given rating) then goto user-item rating matrix else goto step V
- IV) Recommendation list on the basis of highest visit count
- V) end

Step-VI: Weight computation is one in which the score of a recommended product is computed from the results of all the available recommendation techniques present in these approaches. This integrates the scores from each technique using a linear formula. Therefore, the various techniques must be able to produce their recommendation score, which can be linearly combinable. It is very useful that all of the system's capabilities are brought to bear on the recommendation process. The benefit of a weighted hybrid is that all the recommender system's strengths are utilized during the recommendation process in a straightforward way. Hence the ratings of several recommendation techniques are combined together to produce a single recommendation.

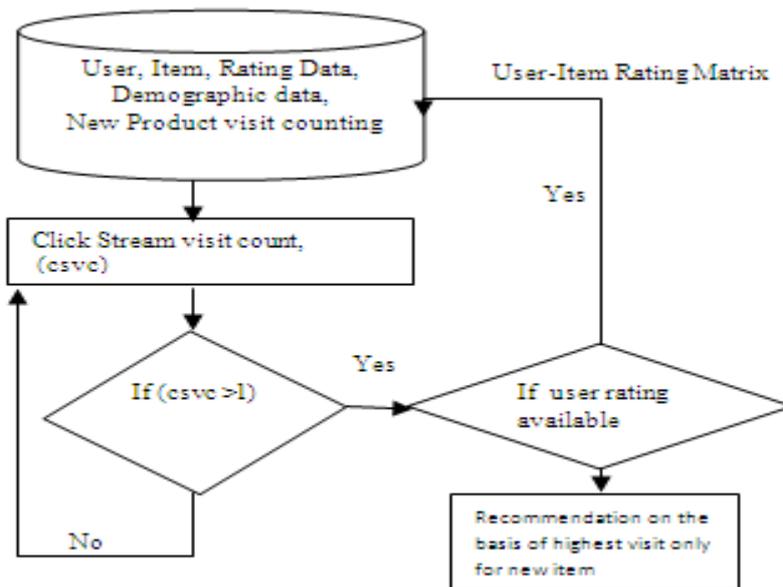


Fig.-2 Flowchart of Click stream visit count

4. CONCLUSION:

This paper proposed the concept of using collaborative filtering, demographic approach, and clickstream visit count. To solve the cold start problem in the recommendation system. This paper mainly introduces the clickstream visit count algorithm along with collaborative filtering and demographic approach. The prediction accuracy of every single recommendation algorithm is insufficient and each has its own merit and demerit, so we pay more attention to hybrid recommendation algorithms. In the weighted strategy, we need to decide how to the optimization of the weight of each recommendation algorithm. Hence the ratings of several recommendation techniques are combined together to produce a single recommendation.

5. FUTURE WORK:

The proposed work is currently on conceptually based. In the near future, we will develop in the experimental environment with real-world data and applications.

REFERENCES:

- [1] F. Ricci, L. Rokach & B. Shapira, "Introduction to Recommender Systems Handbook", in Recommender Systems Handbook 1st ed. USA: Springer, 2011, Ch. 1pp-1-3-5.
- [2] R. Burke, "Hybrid web recommender systems", In the Adaptive web: methods & strategies of web personalization, p. brasilovsky, A. Kobsa & W. Nejdl. Springer 2007, pp. 377-408.
- [3] Schafer JB, Frankowski D, Herlocker J. Sen S, "Collaborative Filtering recommender systems", 2007, the adaptive web Springer Berlin Heidelberg, p.291-324.
- [4] Charu C. Aggarwal, "Recommender Systems" the textbook, 2016, ISBH 978-3-319-29659-3, ©springer, International Publishing Switzerland.
- [5] S. Natarajan, S. Vairavasundaram, A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data", © 2020 published by Elsevier Ltd.
- [6] Park S.T. and Chu W, "Pairwise preference regression for cold start recommendation", 2009, In proceedings of third ACM conference on recommender systems.
- [7] D. Almazro, G. Shahatah, L. Alabdulkarim, M. Khrees, R. Martinez, W. Nzoukou, A Survey Paper on Recommender Systems, 2010, arXiv:1006.5278.
- [8] Qing Li, Byeong Man Kim, "Clustering Approach for hybrid recommender on web Intelligence", © 2003 IEEE.
- [9] A. Said, T. Plumbaum, W.E. De Luca and S. Albayrak, "A comparison of how demographic data affects recommendation", 2011. in Proc. 19th international conference on User modeling, adaption, and personalization,
- [10] D. Zhang, Q. Zou, H. Xiong, CRUC, "Cold-start Recommendations Using Collaborative Filtering in Internet of Things", 2013, arXiv:1306.0165.
- [11] Braunhofer, M., Elahi, M., Ricci, F., "User personality and the new user problem in a context-aware point of interest recommender system." 2015, In: Information and communication Technologies in Tourism. Springer, PP.537-549.

- [12] H. Sobhanam and A. K. Mariappan, "Addressing cold start problem in recommender systems using association rules and clustering technique," 2013 International Conference on Computer Communication and Informatics, Coimbatore, 2013, pp. 1-5.
- [13] Jindal, Honey and Sandeep Kumar Singh. "A Hybrid Recommendation System for Cold-start Problem Using Online Commercial Dataset." , 2014, International Journal of Computer Engineering and Applications, Volume VII, Issue I, July 14.
- [14] N. K. Mishra, V. Mishra and S. Chaturvedi, "Solving cold start problem using MBA," 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, 2017, pp. 1598-1601.
- [15] Xuan Nhat Lam, Thuc Vu, Trong Duc Le, and Anh Duc Duong, 2008, "Addressing cold-start problem in recommendation systems", In Proceedings of the 2nd international conference on Ubiquitous information management and communication (ICUIMC '08). ACM, New York, NY, USA, 208-211.
- [16] Panda ku. Sanjaya, Bhoi ku Sourav, Singh Munesh, "A collaborative filtering recommendation algorithm based on normalization approach", © Springer-Verlag GmbH Germany, part of Springer Nature 2020.
- [17] Lin Guo, Qinke Peng ,” A Combinative Similarity Computing Measure for Collaborative Filtering”, Applied Mechanics and Materials Vols 347-350 (2013) pp 2919-2925 Online: 2013-08-08, © 2013 Trans Tech Publications, Switzerland.
- [18] T.Srivastava, P. Desikan and V.kumar, "Web mining concepts, applications and research directions."2005, In foundations and advances in data mining, Springer, Berlin, Heidelberg, pp. 275-307.

