

Survey on Parallel Hybrid Multigroup Co-clustering Using Collaborative Filtering Model

Mr. Pramod Kale¹, Prof. M. R. Patil²

Dept. of Computer Engineering, SKNCOE, Savitribai Phule Pune University, Pune, India^{1,2}

Abstract: Recommendation systems play a vital role in filtering and modifying the preferred information. Recommender systems are classified into three categories such as collaborative filtering, content-based filtering and hybrid filtering. Generally Collaborative filtering is a technique which typically utilized to construct personalized recommendations on the internet but it suffers from data sparsity and scalability problem. Few websites that uses the collaborative filtering technology include Amazon, Netflix, iTunes, IMDB. The majority of the clustering-based Collaborative Filtering model utilizes only historical rating information in the clustering procedure but disregard other data resources in recommender systems such as the social connections of users and the correlations between items. This paper presents a survey of existing techniques with the novelties highlighting the need of personalized recommendation techniques based on clustering and collaborative filtering. In proposed system we generate recommendations in an effective manner with comparatively better accuracy and least cost by using parallel and distributed approaches with the heterogeneous information sources to overcome the problems like data sparsity and scalability which are very common in recommender systems.

Keywords: Recommender systems, collaborative filtering, co-clustering, information fusion, data sparsity.

I. INTRODUCTION

The World Wide Web development offers a technique for accessing digital information in different domains. Collaborative filtering is one of the most popular and common technique which is used to develop recommender systems. From large amount of available information, which comes from heterogeneous sources, it is very difficult for the users to retrieve or acquire information according to their requirements and preferences. A common and powerful solution tools are used Recommender systems. In day by day life, people depends upon recommendations from other people such as spoken words, reference letters, news reports from news media, general surveys etc, so recommendations plays an significant role in discovering the best items. The information and e-commerce ecosystem are important part of the Recommender systems.

Recommender systems is part of specific type of information filtering system method that helps the peoples to find out their potential interest items by filtering uninterested ones. RS helps the peoples to find out or recommend information items like movies, TV program, music, videos, books, news, images, web pages, scientific literature or social elements such as people, events or groups that are likely to be of interest to the user.

At present, the collaborative filtering recommendation is successful algorithm in RS but it suffers from an two important issues like data sparsity and scalability. First one is accuracy is the so called data sparseness. In addition to data sparseness, there are still many factors influencing the accuracy. Second one is the scalability problem which is really common in real world applications which occurs because of the increasing number of the users and items. To alleviate these problems in this paper we provide an concise survey of the existing research on the field of

recommender systems and gives the categorizing framework that groups some existing techniques which are used to apply in recommender systems to overcoming these two problems. In this survey we also compare their advantage and drawbacks from theoretical point of view. In the rest of this paper is organized as follows, Section II provides a literature survey on recommender systems. Section III describes the proposed systems. In section IV we represent architecture of our proposed systems. Section V we conclude the paper and discuss the future scope.

II. RELATED WORK

In paper [2], author proposes a novel algorithm for recommendation on music by using hyper graph (MRH) and designs the recommendation problem as a ranking problem on a unified hyper graph. Proposed technique utilizes both multiple types of social media information and music acoustic-based content. In proposed system author incorporate multi-source media information by utilizing social media information and music acoustic signals, in music recommendation to improve the performance. Hypergraph generated by using MRH to represents the multitype objects in a music social community as vertices, and relations between these objects as hyper edges. Proposed algorithm significantly provides better results than traditional recommendation algorithms and the social media information is very helpful for music recommendation.

In paper [3], author proposed a personalized recommendation framework that combines the user clustering technology and item clustering technology to solve scalability and data sparsity problems in the collaborative filtering. Based upon users' ratings records on items, user clustering formed and each users cluster has

a cluster center. As per the similarity between targeted user and cluster centers the nearest neighbors of target user can be found and smooth the prediction where necessary. The proposed approach uses the technique of item clustering collaborative filtering to generate the top n recommendations. The recommendation fusion user clustering and item clustering collaborative filtering is more scalable and more accurate.

In paper [4], author proposed a SocialMF model based approach for generating recommendation in social networks which incorporates trust propagation in the matrix factorization approach. Proposed approach reduced recommendation error especially for the problem of cold start users. The feature vector of particular user is dependent on the feature vectors of his direct neighbors in the social network. If a cold start user is not linked to the social network, then this approach no additional information to improve the quality of recommendation for that user.

In paper [5], author proposed a collaborative filtering approach which is based on a weighted co-clustering algorithm extremely useful for a number of online activities such as e-commerce. The key idea is to simultaneously establish user and item neighborhoods through Co-clustering and produce predictions based on the average ratings of the co-clusters. Author presented a new dynamic collaborative filtering scheme based on simultaneous clustering of items and users. Proposed system is scalable for large datasets containing millions of users and items.

In paper [6], author proposed an incorporated construction of social tagging systems, which captures all types of co-occurrence information appearing in tagging data. Proposed social tagging structure is based on author invent a novel matrix-based user profiling scheme to make use of available information in tagging data. The joint real item-tag recommendation very sparse and may subject to noise.

In paper [7], author overviewed the trend and progress of clustering algorithms. This paper presents the solutions to deal with big data challenges by addressing very first proposed algorithms until today's novel solutions. This paper presents multiple algorithms and the targeted challenges for producing better clustering algorithms and then possible future path for more advanced algorithms is discarded based on today's available technologies and frameworks. Although parallel clustering technique is very useful for clustering, but the complexity of implementing such algorithms is a challenge.

In paper [8], author survey on Parallel and Distributed collaborative filtering implementations, aiming not only to provide a comprehensive presentation of the field's development, but also to offer future research orientation by highlighting the issues that need to be further developed. Now a day's great importance is turn towards parallel and distributed

III. PROPOSED WORK

The quality of a recommender system's output is highly depended on the quantity of used data. The more data is available in a recommender systems, the better will be the recommendation.

Table 1: Survey table

Sr. no	Paper	Proposed	Findings
1	Music recommendation by unified hyper graph: combining social media information and music content [2]	Author proposed a novel algorithm for music recommendation on hyper graph (MRH) and designs the recommendation problem as a ranking problem on a unified hyper graph.	Proposed algorithm significantly provides better results than traditional recommendation algorithms and the rich social media information is very helpful for music recommendation. Performance is improved by proposed system.
2	A collaborative filtering recommendation algorithm based on user clustering and item clustering [3]	Author proposed a personalized recommendation approach that combines the user clustering technology and item clustering technology to solve scalability and sparsity in the collaborative filtering	The recommendation fusion user clustering and item clustering collaborative filtering is more scalable and more accurate.
3	A matrix factorization technique with trust propagation for recommendation in social networks [4]	Author proposed a novel Social MF model based approach for recommendation in social networks which incorporates trust propagation in the matrix factorization approach.	Proposed approach achieves significantly reduced recommendation error (RMSE) in particular for cold start users.
4	A scalable collaborative filtering framework based on co-clustering. [5]	Author proposed a novel collaborative filtering approach based on a weighted co-clustering algorithm	Proposed system is scalable for large datasets containing millions of users and items.
5	Big Data Clustering: A Review(7)	This paper introduces concepts and algorithms related to clustering, a concise survey of existing (clustering) algorithms as well as providing a comparison, both from a theoretical and an empirical perspective.	This survey provided a comprehensive study of the clustering algorithms proposed in the literature. In order to reveal future directions for developing new algorithms and to guide the selection of algorithms for big data,

Having to deal with continuously growing amounts of data, the design of parallel and distributed recommender systems has become necessary. The parallel and distributed computing techniques can be combined with each other to the purpose of exploiting their advantages and various modifications can be applied to the existing algorithms in order to fit better to the requirements of the used techniques. Furthermore, taking advantage of the heterogeneous infrastructures that are available is crucial for the development of high quality recommender systems. In our system we propose a hybrid recommendation approach based on information fusion. We utilize the information from heterogeneous sources to overcome the problem of data sparsity. Besides rating matrix we also use the users social network's and items co-relations information. Traditional clustering algorithms are fail to generate accurate recommendation because of their assumption is that each user or item can belongs to particular group. With the information from the first module we co-cluster users and items in multiple groups to overcome the above mentioned problem.

We also use the parallel clustering approach to overcome the problem of scalability. By using parallel clustering technique the system is capable to handle large amount of users and items and produces the faster results which improves the performance of the system.

IV. ARCHITECTURAL VIEW

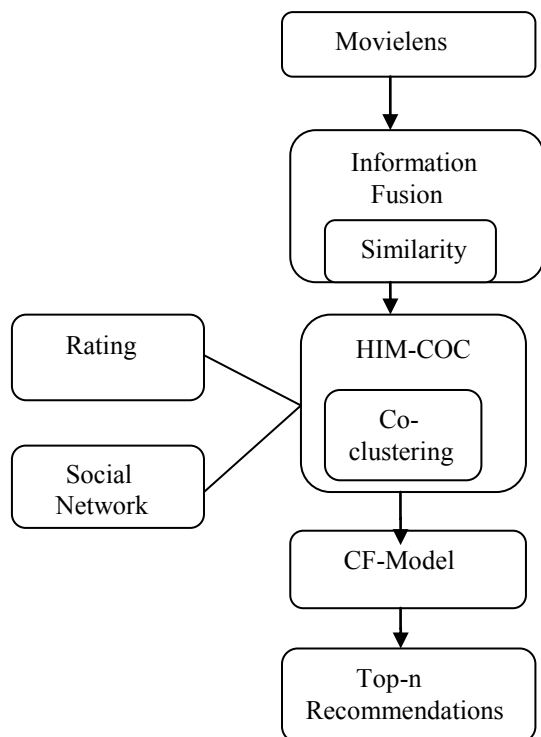


Figure1: System Architecture

V. CONCLUSION

This paper presented an all-inclusive survey on recommendation techniques based on clustering and collaborative filtering. The main features, the advantages

and disadvantages of each in recommendation system are described. The most important issue which influences the collaborative filtering recommendation accuracy is the so-called data sparseness. As per survey, there is need of novel approach to generate recommendations in an effective manner with comparatively better accuracy and least cost. Thus, the future scope is to study and design of parallel algorithms and implementations that will address the emerged problems and exploit the advantages of new technologies is important.

REFERENCES

- [1] Huang, Shanshan, et al. "A Hybrid Multigroup Coclustering Recommendation Framework Based on Information Fusion." *ACM Transactions on Intelligent Systems and Technology (TIST)* 6.2 (2015): 27.
- [2] Bu, Jiajun, et al. "Music recommendation by unified hyper graph: combining social media information and music content." *Proceedings of the international conference on Multimedia*. ACM, 2010.
- [3] Gong, Songjie. "A collaborative filtering recommendation algorithm based on user clustering and item clustering." *Journal of Software* 5.7 (2010): 745-752.
- [4] Jamali, Mohsen, and Martin Ester. "A matrix factorization technique with trust propagation for recommendation in social networks." *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 2010.
- [5] George, Thomas, and Srujana Merugu. "A scalable collaborative filtering framework based on co-clustering." *Data Mining, Fifth IEEE International Conference on*. IEEE, 2005.
- [6] Peng, Jing, et al. "Collaborative filtering in social tagging systems based on joint item-tag recommendations." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010..
- [7] Shirikhshidi, Ali Seyed, et al. "Big data clustering: A review." *Computational Science and Its Applications-ICCSA 2014*. Springer International Publishing, 2014. 707-720.
- [8] Karydi, Efthalia, and Konstantinos G. Margaritis. "Parallel and Distributed Collaborative Filtering: A Survey." *arXiv preprint arXiv:1409.2762* (2014).
- [9] Dhillon, Inderjit S. "Co-clustering documents and words using bipartite spectral graph partitioning." *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001.
- [10] Di Noia, Tommaso, et al. "Exploiting the web of data in model-based recommender systems." *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 2012.