Abhishek Raj Permani, 2022, 10:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

Community Detection using Girvan Newman algorithm in Recommendation systems

Abhishek Raj Permani, Swapnil Narayan
Department of Computer Science and Engineering,
IIT (ISM),
Dhanbad, Jharkhand, India,
abhishekraj29011998@gmail.com, swapism7@gmail.com

Abstract- Any e-commerce site we visit, there is the set of recommendations for us already, it might be based on many parameters like- our search history, shopping preferences, location we are located and many more. Collaborative filtering techniques are popularly used techniques for giving recommendations to the users. But Collaborative technique also suffers with cold start problem, when some new item or user is operated on the system. This paper approaches this problem using hybrid techniques of Girvan Newman algorithm-based alternating least square factorization. Girvan Newman algorithm is a hierarchical method used to detect communities in complex systems and alternating least square algorithm is used to predict recommendations.

Keywords- **Recommendation System, Community Detection, Girvan Newman, Prediction**

I. INTRODUCTION

Daily surfing of e-commerce website generates a lot of data containing ratings, reviews, opinions of many items present there. More the information that is more the data available for any item, better the recommendation system working will be for that item.As the social network data is a big data, the efficiency and scalability requirements can be addressed using map reduce techniques with Hadoop [1, 3, 5, 8]. Weak ties theory introduced two aspects such as homophily and heterophily to find the interactions between the online social network users [2, 6, 9, 10].

There are many Recommendation Systems based on various Collaborative filtering techniques like- user based and item based mostly. User based technique measures the similarity between target and other users, on the other hand item-based techniques measures similarity between items that target users' rate or interact with and other items. Considering the example of some big e-commerce giants like Amazon, where Amazon is known for its utilization of cooperative separating, coordinating with items to clients dependent on past buys. For instance, the framework can distinguish every one of the items a client and clients with comparative practices have bought or potentially decidedly evaluated.

The similarity measures are identified by using multi-feature vector as user preferences and contexts, which are combined by memory-based approach. In the iterative approximation steps, the contexts and preferences are integrated to learn the procedures for item ratings by model-based methods.

This research work introduces a new hybrid social Recommendation System Girvan Newman algorithm-alternating least square algorithm for giving recommendations where user–user social graph is analyzed by identifying the communities from the user interactions using Girvan Newman algorithm, user–item ratings are identified by using user–item based Collaborative filtering and finally recommendations are generated using alternating least square algorithm.

II. LITERATURE REVIEW

The audit of late strategies is talked about in this part which depends on different sorts of Community Filtering. The procedures are clarified exhaustively with their benefits and constraints.

A social Recommendation System using Community Detection and Community filtering as a community-wise social interactions technique was developed by Lalwani et al. [4].

The user–user social graph was analyzed to extract the relationship between friendships by using Community Detection algorithm. The Community Filtering method was used for rating prediction, based on user–item. The MapReduce framework was developed to present a community based social Recommendation System and scalable Community Filtering.

The coverage rate and scalability were improved by using these methods when compared with traditional algorithms. In addition, the issues of Cold start were effectively handled by this approach but failed to concentrate on data sparsity issues.

Park et al. [7] developed a Reversed Community Filtering (RCF) to provide better recommendation, where the processing time of Reversed Community Filtering (RCF) was decreased significantly by using k-nearest neighbors (KNN) algorithm as greedy filtering.

III. BACKGROUND

The section contains the general architecture of social recommendation system, the problems of CS, background of Girvan Newman algorithm to detect the community and alternating least square are discussed.

1. Architecture of social recommendation system:
Social proposal frameworks assume a significant part to give proposals or to offer ideas to the clients, who are from online informal communities like Facebook, Twitter, and Google and so on the ideas or suggestions are given dependent on client inclinations, interests and collaborations among them. For the most part friendly suggestion framework will attempt to take care of the basic issues like information sparsity or Cold State client issues while giving proposals. Taking care of these kinds of issues is actually a difficult undertaking for suggestion frameworks.

Consequently, there should be a legitimate way to deal with give proposals by taking care of these issues. The overall design of social proposal framework is displayed in the Fig. 1.

Bolts show the progression of information to the separate stages. At first client's data like their profile, connections with their companions, relationship with their organizations and so on are gathered from online informal communities [5, 28]. Later the data is taken care of to the proposal motor which comprises of three stages in particular local area discovery, suggestion model and forecast model.

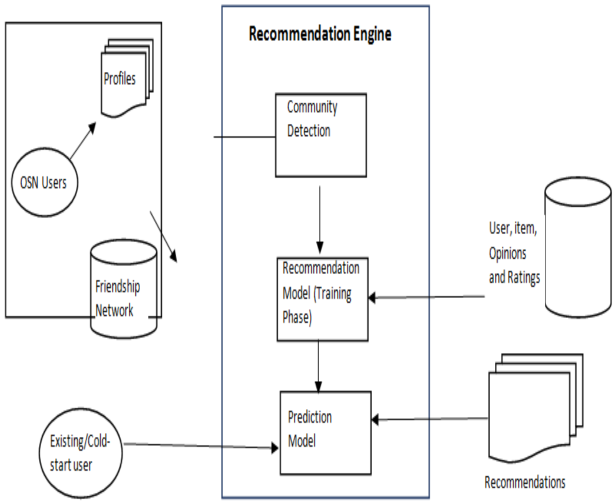

Fig 1. Architecture of Social recommendation system.

2. Community Detection:
Community Detection stage distinguishes the networks from online interpersonal organization clients by considering client interests, conclusions and so forth A ton of Community Detection calculations are accessible identifying client networks. One of the well-known calculations is Louvain's calculation.

3. Recommendation model (training):
As the information is tremendous for handling, there ought to be a legitimate proposal model to prepare the information. Information sparsity issue is tended to during the preparation stage during which every one of the missing passages for Cold Start clients are filled. Preparing stage might utilize AI or neural organizations or enormous information methods for handling tremendous information. Number of calculations is accessible for preparing reason. ALS is one of those calculations.

4. Prediction model:
The resultant prepared information from suggestion model is utilized for expectation. There are number of approaches for anticipating the last suggestions. Some of them are, taking the top k suggestions by discovering normal rating networks or taking the proposal from client thing planning lattice and so forth Determination of expectation model is finished by the application.

5. Issues in cold start:

Cold Start client is another client to the proposal framework who has never been the piece of online informal community. By and large, if a suggestion is to be given to client by the recommender framework, he ought to be a piece of online informal organization. Yet, there is no data having a place with Cold Start client in the web-based informal organization as he is new. Because of the absence of data, it is extremely challenging to offer proposals to the CS client. This issue can be redressed by utilizing legitimate strategies in the proposal frameworks while building the suggestion model.

6. Girvan Newman's community detection algorithm:

There are numerous web-based informal community clients in the web. To concentrate on the conduct of online informal community, there is a need to unite the practices of clients as there will be a social match between the clients. Thus, comparative clients should be assembled into each gathering to give combination results. These gatherings are called as networks. Better union outcomes can be accomplished by successfully making networks.

We have used a popular graph partitioning algorithm used for community detection, i.e. Girvan Newman Algorithm. The Girvan Newman Algorithm removes the edges with the highest between's until there are no edges remain. Between's the number of the shortest paths between pairs of nodes that run through it.

7. Alternating least square:

One of the proficient grid factorization methods in the RSs is ALS procedure. Generally, it is utilized to anticipate suggestions by tracking down the missing passages in the given framework. Here the network is client thing audit evaluations lattice. While computing the upsides of ALS, the minimization rule esteem is fixed, on the grounds that that factorization isn't raised. In this work, we created client thing rating (A × B) lattice for every local area in G where An addresses clients and B addresses films and the opening relating to An and B addresses rating. Presently, we apply the network factorization utilizing ALS to A × B lattice to create introductory irregular grid X × Y where X addresses client interests' components and Y addresses film types of factors. Therefore, the client thing rating network is determined as introduced here.

$$X = \begin{bmatrix} | & & | \\ x_1 & \dots & x_n \\ | & & | \end{bmatrix}, \quad Y = \begin{bmatrix} | & & | \\ y_1 & \dots & y_n \\ | & & | \end{bmatrix}$$

To get vectors xu, the X matrix (user reviews) are calculated by fixing the Y matrix (movie ratings), and vice versa. A quadratic function problem is created, when this procedure iterates until converges. Initially, while fixing the Y, the equation over X are reduced and this function becomes as in Eq. (1)

$$L(X) = min_{X,Y} \sum_{ui \in \Omega} \left(r_{ui} - x_u^T y_i\right)^2 + \lambda \sum_u |x_u|^2 + \lambda \sum_i |y_i|^2 \quad (1)$$

where R's nonzero ratings are defined as Ω , and uth row vectors of matrix X is represented as xT u, ith column of matrix Y is explained as yi, and to avoid the over-fitting problem, regularized coefficient are considered as constant as λ. By calculating the partial derivative of xu and yu in Eq. (4) and equating to zero, then the Eq. (4) is modified as given in Eq. (2),

$$x_u = \left(Y^T Y + \lambda I\right)^{-1} Y r_u \quad (2)$$

Where unit matrix ranked f is described as I, and ru is the uth row of R. The value for yi are achieved in the same way, which are given in Eq. (3)

$$y_i = \left(X^T X + \lambda I\right)^{-1} X r_i \quad (3)$$

Once the X and Y matrix values are finalized, then they can be used for predicting recommendations.

IV. PROPOSED APPROACH

The square chart of proposed GN-ALS procedure is shown the Fig. 2. At first client informational index is gathered from the Online Social Network (OSN) which is considered as contribution to Girvan Newman. Later client networks are made by utilizing Girvan Newman's calculation as clarified in the Sect. 3.3.

It begins with every client locally and afterward the local area is augmented by adding more clients with comparative inclinations at every cycle. At one phase, wanted measured quality is accomplished, and no further blending occurs. The calculation then, at that point, delivers a concluded arrangement of networks. Client thing rating networks are created for the finished networks as per their connections or mappings in the given informational index. These networks may accompany the issue called information sparsity issue.

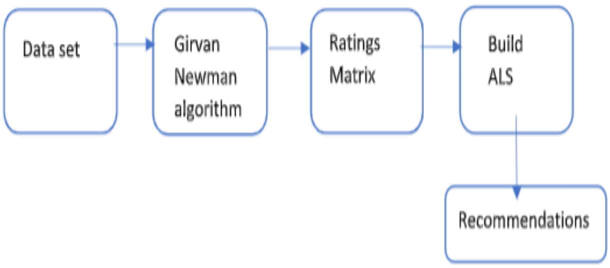

Fig 2. General block diagram of GN-ALS methodology.

Information sparsity alludes to the trouble in discovering enough solid comparable clients where dynamic clients just rate a little part of things. In this work, information sparsity issue is tended to by utilizing grid factorization procedure called ALS approach with the assistance of Girvan Newman's calculation. Tables 6 and 7 in the paper presents the procedure took on utilizing ALS calculation.

ALS calculation produces client thing lattice from the dataset and limits sparsity via preparing the model and properly filling the clear qualities in the grid. The interaction begins with task of irregular qualities and the blunder is limited in an iterative way with the quantity of emphases.

The utilization of ALS calculation brings about limiting the mistake between real qualities and qualities inferred by ALS for filling the fragmented information. The mistake is determined, and values are refreshed until every one of the qualities is united. The ALS calculation is prepared to construct a proposal model. It takes the networks and rating grids produced from Girvan Newman's calculation to determine the forecast model.

When the expectation model is acquired, another client/leaving client information can be given as contribution to get the anticipated proposals. The ALS calculation fills in as displayed in the Sect. 3.4. The undertaking of ALS is to anticipate the evaluations given by the client for motion pictures and afterward prescribe that to client as per appraisals. The proposed GN-ALS algorithm is shown in below:

Algorithm GN-ALS(g,n,I)

{

// g represents online users social network

// n represents number of users

// i represents movie item ratings given by users

Input: Initialize the network g with n number of nodes

// each node i represents a user in the network g

Output: user-item rating matrix after matrix factorization

Start:

1. Calculate modularity score for g
2. Calculate the between's of all existed edges in graph.
3. Now remove all the edges(s) with the highest between's.
4. Now recalculate the between's of all the edges that got affected by the removal of edges.
5. Now repeat steps 3 and 4 until no edges remain.
6. Create a new network by grouping all of the nodes in the same community.
7. Now each community is treated as a newly formed node in the newly created network G
8. Assign new network G to step until there is no change in G.
9. Generate user-item rating (A x B) matrix for each community in G where A represents users and B represents movies and the slot corresponding to A and B represents rating.
10. Apply matrix factorization problem using GN to A x B matrix and generate initial random matrix X x Y where X represents user interests factors and Y represents movie types factors.
11. Lets us assume there are m rows in X and n columns in Y.
12. For iteration: =1to Max

    {

13. For row u←1 to m

    {

    1. $x_u \leftarrow (Y^T Y + \lambda I)^{-1} Y r_u$
    2. //r_u represents u^{th} user row ratings data.

    }// end for

14. For column i← 1 to n

    {

1. $y_i \leftarrow (X^T X + \lambda I)^{-1} X r_i$
2. // r_i represents i^{th} column ratings data

}
//end for
}
//end for

15. Now consider the final X x Y matrix for recommendation
16. Stop

}
// end algorithm

The resultant X × Y framework is utilized for giving proposals. To get proposals, the CS client is given as contribution to proposed technique. Later people group of the CS client is distinguished. When his local area is distinguished, ALS calculation is applied, and all film thing appraisals will be taken from definite X × Y framework for that client local area. Presently the top K film things are prescribed to the CS client which are first class among all the film thing appraisals taken from the before.

V. CONCLUSION

Social recommendation systems are very much important to give recommendations to the online users. Cold Start problems arise when the users are new to social network. The Cold Start problem can be effectively addressed by adopting proper methodologies for getting better recommendations. The proposed Girvan Newman algorithm-based alternating least square factorization approach is the one such methodology for addressing Cold Start problem effectively.

REFERENCES

[1] Balaga TR, Peram SR, Paleti L (2017) Hadoop techniques for concise investigation of big data in multi-format data sets. In: 2017 2nd international conference on communication and electronics systems (ICCES). IEEE, pp 490–495.
[2] Granovetter MS (1977) the strength of weak ties. In: Leinhardt S(ed) Social networks. Elsevier, Amsterdam, pp 347–367
[3] Jin S, Lin W, Yin H, Yang S, Li A, Deng B (2015) Community structure mining in big data social media networks with mapreduce. Clust Comput 18(3):999–1010
[4] Lalwani D, Somayajulu DV, Krishna PR (2015) A community driven social recommendation system. In: 2015 IEEE international conference on Big Data (Big Data). IEEE, pp 821–826.
[5] Lee WP, Tseng GY (2016) Incorporating contextual information and collaborative filtering methods for multimedia recommendation in a mobile environment. Multimed Tools Appl75(24):16719–16739
[6] McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27(1):415–444
[7] Park Y, Park S, Jung W, Lee SG (2015) Reversed cf: A fast collaborative filtering algorithm using a k-nearest neighbor graph. Expert SystAppl 42(8):4022–4028.
[8] Wang Y, Wang M, Xu W (2018) A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics frame work. Wirel Commun Mobile Comput. https: //doi.org/10.1155/2018/82637 04
[9] Yang C, Lan S, Shen W, Huang GQ, Wang X, Lin T (2017) Towards product customization and personalization in IoT-enabled cloud manufacturing. Clust Comput 20(2):1717–1730
[10] Yuan T, Cheng J, Zhang X, Liu Q, Lu H (2015) How friends affect user behaviors? An exploration of social relation analysis for recommendation.