Solving Sparse Rating Problem Using Fine Grained Approach

Nibha Sharma
Student, Computer Science
IIIT Bhubaneswar
Bhubaneswar, India
A110013@iiit-bh.ac.in

Mudasir Mohd
Student, Computer Science
IIIT Bhubaneswar
Bhubaneswar, India
A110012@iiit-bh.ac.in

Anjali Mohapatra
Assistant Professor, Computer Science
IIIT Bhubaneswar
Bhubaneswar, India
anjali@iiit-bh.ac.in

***Abstract*— Recommender System is a system that automatically recommends all similar kind of items that are of user interest. In design of the recommender systems rating is the crucial issue. Till today many algorithms have been proposed for efficient recommendation but they still requires further improvements to make it more effective. In this paper we address the limitations of recommendation methods and propose a possible model to address the First rater problem(Sparse Rating) of the collaborative based approach to improve recommendation capabilities for a broader range of applications.**

Keywords-Recommender System; Machine Learning; Collaborative Filtering

I. INTRODUCTION

Information available over web is increasing day by day which create inconveniency for users to access the data efficiently. In this above scenario users and information are directly proportional to each other which gives rise to the need of a system which discovers content of user interest in less effort. For fulfilling this need a system called Recommender System has been introduced which recommends data to the user according to his interest. Recommendation generally reduces the job of the user to search for more similar items he/she looking for. A recommender system automatically recommends all similar kind of items that are of user interest. Once a user rate for some item, then from next time onwards all the similar items will be recommended to him according to the rating given by him in the past. Examples include recommending books, CDs and other products at Amazon.com, movies by Movie Lens, and news at VERSIFI Technologies (formerly AdaptiveInfo.com). Moreover, some of the vendors have incorporated recommendation capabilities into their commerce servers. In its most common formulation, the recommendation problem is reduced to the problem of estimating ratings for the items that have not been seen by a user. Intuitively, this estimation is usually based on the ratings given by the user to other items and on some other information. Once we will be able to estimate ratings for the yet unrated items, we can recommend the item(s) with the highest estimated rating(s) to the user. However, despite many advances, the recommender system still requires further improvements to make recommendation methods more effective.

In this paper, we will discuss the existing recommender system and will identify some limitations related to collaborative recommenders which can create problem in recommendation and propose a model to address this. The rest of this paper is organized as: a comprehensive survey of existing recommender systems along with the

limitations are presented in Section 2. Our proposed approach for a possible solution to solve some of these issues is presented in Section 3 followed by conclusion and future works in Section 4.

II. EXISTING RECOMMENDER SYSTEMS

Recommender system emerged as an independent research area in mid-1990's when researchers started focusing on recommendation problems that explicitly rely on the ratings structure. There are many recommender systems that are working successfully in different fields such as in cognitive science, approximation theory, information retrieval, forecasting theories, and also have links to management science, to the consumer choice modeling in marketing etc. These usually work over the user rating or user preferences. Based on how recommendations are made, these systems are usually classified into Content-based recommendations, Collaborative recommendations, and Hybrid approaches. In Content-based recommendations, the user is recommended items similar to the ones the user preferred in the past where as in Collaborative recommendations, the user is recommended items that people with similar tastes and preferences liked in the past. Hybrid approaches combine collaborative and content-based methods depending on the nature of application. In this paper we will focus on the problem of collaborative recommenders.

A. Issues Related to Recommender Systems

In this section we will discuss the fundamental problems that Recommender Systems suffer from.

1. Recommendation Systems based on Collaborative-based Approach encounters the following drawbacks:
   - Sparsity: It is usual in e-business stores that even the most active customers have purchased or rated a very limited percentage of products, when compared to the available total. That leads to sparse user-item matrices, inability to locate successful neighbors and finally, the generation of weak recommendations. As a result, techniques to reduce the sparsity in user-item matrices should be proposed [1] [2].
   - Unusual User Problem: It is also known as the Gray Sheep problem. It refers to individuals with opinions that are "unusual", meaning that they do not agree or disagree consistently with any group of people. Those individuals would not easily benefit from recommender systems since they would rarely, if ever, receive accurate predictions [1].
   - Scalability: Recommender Systems require calculations, which grow with both the number of customers and the number of products. An algorithm, that is efficient when the number of data is limited, can turn out to be unable to generate a satisfactory number of recommendations, as soon as the amount of data is increased. Thus, it is crucial to apply algorithms, which are capable of scaling up in a successful manner [2].
   - First Rater Problem: An item cannot be recommended unless a user has rated it before. This problem applies to new items and also to obscure items and is particularly harmful to users with eclectic tastes. It is also known as the Cold Start Problem [3] [4].
2. Recommendation Systems based on Content-based Approach encounters the following drawbacks:
   - New user problem. The user has to rate a sufficient number of items before a content-based recommender system can really understand user's preferences and present the user with reliable recommendations. Therefore, a new user, having very few ratings, would not be able to get accurate recommendations.
3. Other Issues related to recommendation systems:
   - Quality of Recommendations: Trust is the key word here. Customers need recommendations, which they can trust. To achieve that, a recommender system should minimize the false positive errors, i.e. products, which are recommended (positive), though the customer does not like them (false).Such errors lead to angry customers [5].

In this paper we will focus on First Rater Problem of the Collaborative-based Approach and propose a model to address this by combining approaches of content-based and collaborative based recommender system.

III. PROPOSED MODEL

When a user is searching for some item for the first time, its rating is zero. First Rater Problem(Sparse Rating) reduces the performance of a Recommendation System as with zero rating ,recommendation system cannot

recommend anything to the user. So in order to address this issue a Hybrid approach is proposed that uses both the techniques, content-based and collaborative.

A. Method

The proposed approach comprises of following components:

1) Topic modeler.
2) Indexer.
3) Rating generator
4) Recommending System

Topic Modeler: Topic Modeler consists of Topic Modeling tool. MALLET is a topic modeling tool which is referred for our analysis purpose. We have chosen LDA as our topic model. In this, we provide a set of documents and it returns a set of topics for the documents according to the probability distribution of the topics and then we choose topic with best probability out of it.[6][7]

Indexer: It is a tool that index documents on the basis of their topics and returns a chained structure which has topic and all documents related to that topic.

Rating generator: This will provide rating to the topics and documents which we will use further for recommending items to the user.

Recommending System: Recommender System typically produces a list of recommendations either through collaborative or content-based filtering. Collaborative filtering approaches to build a model from user's past behaviors as well as similar decisions made by other users and use that model to predict items that the user may have an interest in. A content-based filtering approach utilizes a series of discrete characteristics of an item in order to recommend additional items with similar properties. These approaches are often combined called hybrid Recommender Systems.

B. Algorithm

STEP1: First time a user logs in and reads a document, that document will be stored in the list if it is not preexisting in the list and then it is provided to topic modeler (Latent Dirchlet Allocation) which assign most promising topic to this document(say T1).
STEP2: If document is already present in the list then it is assigned to rating calculator which calculates rating of the document.
STEP3: T1 is feed to the Indexer, indexing will be performed over documents, resulting all documents that belongs to T1. This will provide a list of indexed topics.
STEP4: After rating and indexing, both the results are provided to recommender system which maps both the results and recommend most suitable document to the user. This recommendation is based on the assumption that rating of the topic will be the rating of highest rated document and rating of document will be average rating of all documents within the same topic.
STEP5: The topic of the document read by the user is calculated and is moved to the particular index of that topic . The referred index is scanned and the document read by the user is provided with the average rating of all the documents present in the index and the topic is also rated. On the basis of both ratings, the documents are recommended to the user.

Above steps are diagrammatically shown in Fig. 1 , Fig. 2 and Fig. 3.

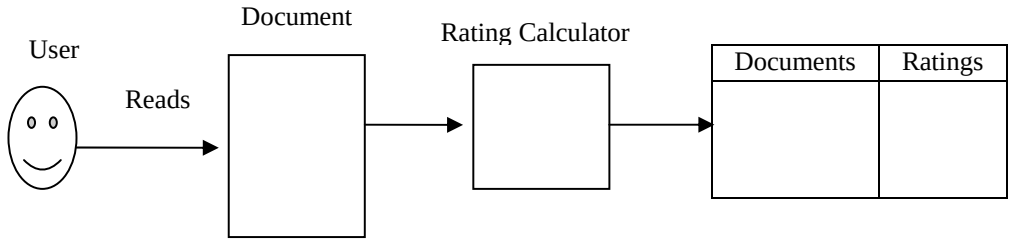

Fig. 1 System A

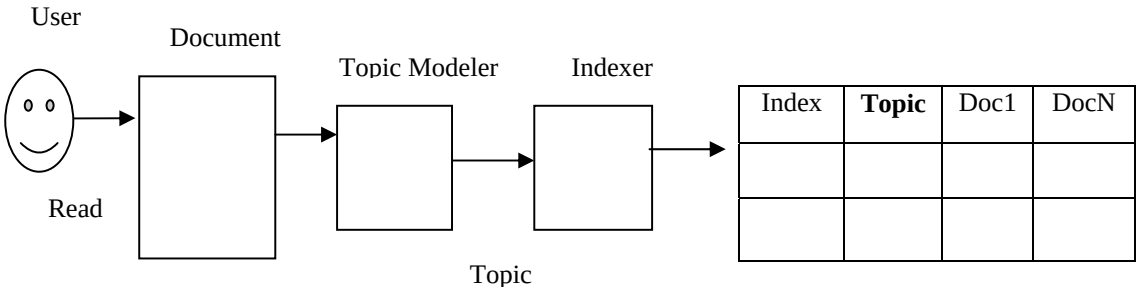

Fig. 2 System B

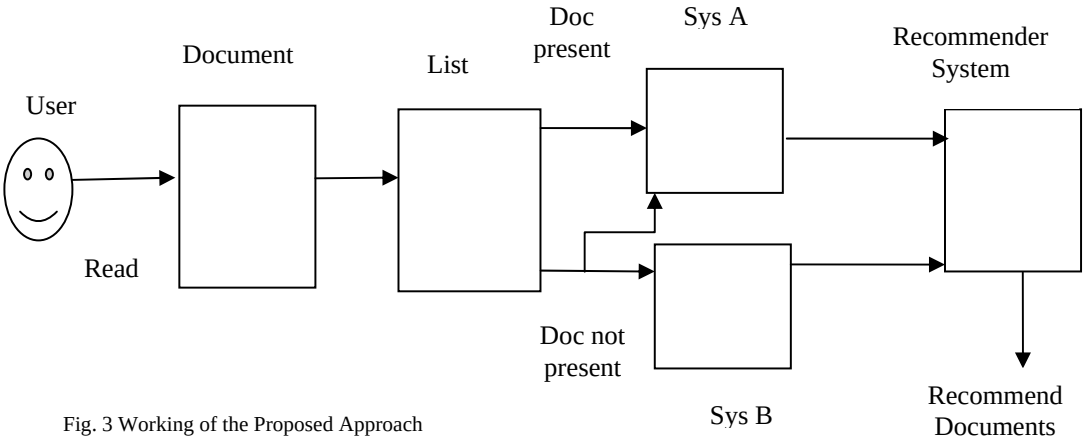

Fig. 3 Working of the Proposed Approach

IV. CONCLUSION AND FUTURE PLANS

In this paper we considered collaborative filtering technique and problem associated with it. Followed by an overview of Sparse Rating problem. We proposed a possible model that can eliminate the Sparse Rating problem of collaborative based approach using a hybrid model and improve the quality of recommendations. In Future work we would like to conduct further experiment using hlda (Hirerichal LDA) to extensively evaluate the proposed method.

REFERENCES

[1] Mark Claypool, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, and Matthew Sartin, "Combining content-based and collaborative filters in an online newspaper," ACM SI-GIR Workshop on Recommender Systems-Implementation and Evaluation,Berkeley,CA: 1999.

[2] Bardul M. Sarwar., "Sparsity, Scalability, and Distribution in Recommender Systems," Ph.D. thesis, University of Minnesota:2001.

[3] Prem Melville, Raymond J. Mooney, and Ramadass Nagarajan, "Content-boosted collaborative filtering ," ACM SIGIR Workshop on Recommender Systems, New Orleans,LA: 2001.

[4] Andrew I. Schein, Alexandrin Popescul, Lyle H. Ungar, and David M. Pennock, " Methods and metrics for cold-start recommendear," ACM SIGIR-2002, Tampere,Finland:2002.
[5] Bardul M. Sarwar, George Karypis, Joseph A. Konstan, and John T. Riedl, "Analysis of recommendation algorithms for e-commerce," ,Electronic Commerce, 2000.
[6] David M. Blei, Andrew Y. Ng and Michael I. Jordan," Latent Dirichlet Allocation ,"Journal of Machine Learning Research 3,2003.
[7] David M. Blei and John D. Lafferty,"Topic Models," 2009

AUTHORS PROFILE

Nibha Sharma received his B.E degree (Computer Engineering) in 2010 from Rajiv Gandhi University, Bhopal, Madhya Pradesh, India. Currently she is pursuing her M.Tech degree from Department of CSE, IIIT Bhubaneswar, Odisha, India. Her area of research is Natural language processing and Data Mining.
E-mail: A110013@iiit-bh.ac.in

Mudasir Mohd received his MCA degree in 2009 from Kashmir University, Srinagar, J&K, India. Currently he is pursuing his M.Tech degree from Department of CSE, IIIT Bhubaneswar, Odisha, India. His area of research is Natural language processing.
E-mail: A110012@iiit-bh.ac.in

Anjali Mohapatra received her Ph.D. degree in 2011 from Utkal University, Bhubaneswar, Orissa, India. Currently she is working as an Assistant Professor in the Department of CSE, IIIT Bhubaneswar, Odisha, India. Her areas of research are Bioinformatics,Algorithms and design,NLP.
E-mail: anjali@iiit-bh.ac.in