Data Analytics Using Hadoop Framework for Effective Recommendation in E-commerce Based on Social Network Knowledge

Khyati P Raval[1] ,Dr. Shayamal Tanna[2]
[1]Student,Department of Computer Engineering,L. J. I. E. T, Ahmedabad, Gujarat, India
[2]Assistant Professor,Department of Computer Engineering,L. J. I. E. T, Ahmedabad, Gujarat, India

Abstract

Nowadays Internet has become the basic source of information. variety of different approaches and algorithms of data filtering and recommendation are available. In traditional recommender system user can not get satisfactory result because many problems occurs during the process for instance changing in user preference problem and cold start problem. So, for making recommender system strong the developers uses different criteria like, extracting user data from social networking sites. Here, we are trying to describe the recommendation system in e-commerce using social network knowledge base and then introduces various techniques and approaches used by the recommender system like, User-based approach, Item based approach, Hybrid recommendation approaches and also shown related research work in the recommender system. In the end we will show the Idea about my proposed system.

Keywords—*Hadoop, Data Analytics, Recommendation system, E-commerce, AHP process, Data sparsity*

.1.INTRODUCTION

The growth of the technology and the big usage of recommendation system in many systems like in learning system, tourism system, and ecommerce system gives focus on the techniques used in those system development. Recommendation systems are defined as a software tool and techniques which providing advice for item to a user. The suggestions are like what music to listen, what online news to read etc. Recommendation system is used for finding the needed information from wider information available on the internet. Recommendation system mainly uses three approaches content based recommendation system, collaborative filtering recommendation system and hybrid recommendation system.[7]

The First recommender system was developed by Gold-berg, Nichols, Oki and Terry in 1992. Tapestry was an electronic messaging system that allowed users to either rate messages (good or bad) Recommender system as defined by M. Desh-pande and G. Karypis: A personalized information filtering technology used to either predict whether a particular user will like a particular item (prediction problem) or to identify a set of N items that will be of interest to a certain user. Recommender systems form or work from a specific type of information filtering system technique that attempts to recommend information items (movies, TV program/show/episode, video on demand, music, books, news, images, web pages, scientific literature etc.) or social elements (e.g. people, events or groups) that are likely to be of interest to the user. Typically, a Recommendation system compares a user profile to some reference characteristics, and seeks to predict the 'rating' or 'preference' that a user would give to an item they had not yet considered. These characteristics may be from the information item (the content-based approach) or the user's social environ-ment (the collaborative filtering). The recommender system apply data mining techniques and prediction algorithms to predict us ers interest on information, product and services user. Recommender systems apply techniques and methodologies from another neighbouring areas - such as Human computer interaction (HCI)or Information Retrieval(IR).However, most of these systems bear in their core an algorithm that can be understand as a particular instance of a data mining(DM) technique. [7]

A. BACKGROUND:- CLASSIFICATION OF RECOMMENDATION TECHNIQUES

1) Collaborative filtering: Collaborative filtering methods are based on collecting and analyzing a large amount of information on users behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an ”understanding” of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems. For example, the k-nearest neighbour (k-NN) approach and the Pearson Correlation as first implemented by Collaborative Filtering is based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past. [8]

2) Content-based filtering: Another common approach when designing recommender systems is content-based filtering. Content-based filtering methods are based on a description of the item and a profile of the users preference. In a content-based recommender system, keywords are used to describe the items; beside, a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in
the past (or is examining in the present). In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research. To abstract the features of the items in the system, an item presentation algorithm is applied. A widely used algorithm is the tf idf representation (also called vector space representation. [8]

3) Hybrid Recommender Systems: Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them; by adding content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model for a complete review of recommender systems). Several studies empirically compare the performance of the hybrid with the pure collaborative and content-based methods and demonstrate that the hybrid methods can provide more accurate recommendations than pure approaches. These methods can also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem. [8]

B. Hadoop:-

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop framework application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

Hadoop provides its own file system called HDFS (Hadoop Distributed File System). When we deploy our text data on Hadoop file system, Hadoop distributes all the data in different clusters and performs operations parallel. Hadoop also keeps multiple copies of data in case of hardware failure. HDFS represents data in the pair of {key, value} associated with specific key. Here we have data in the form of text then each word can identify with a key and its associated value represent its weightage or occurrence in the whole dataset. Applying the concept of map reduce where mapper class work as partitioner and reducer class work as combiner.[6]

2. RELATED WORK

In SNetRS: Social Networking in Recommendation System Jyoti Pareek et al. proposed the recommendation system for ecommerce websites which address issue of cold start problem and change in user preference problem. Proposed SNetRs Recommendation system algorithm is based on collaborative filtering which focuses on the user preferences obtained from social media. They are going to further research on the same topic. They plan to implement this model and to add time factor and cross domain filtering. Time factor model will help in knowing the rating gaps base on time. Cross domain filtering will help to know the purpose of user, visiting our site. From cross domain filtering system will get an idea, about the product user is looking for. [3]

In An Efficient and Optimized Recommendation System Using Social Network Knowledge Base Md Zeeshan Ashraf et al. present a novel recommendation system for e-commerce websites using social network knowledge base that uses certain parameters provided by users viz. age group, gender, location etc. and based on these criteria best recommendation is provided by our proposed method using Analytical Hierarchy Process, Merge-

and-Sort, and Sort-and-Count algorithms within a wrapper to optimize. Finally conclude from our research and analysis that, Recommendation System is here to stay. Social networking being the best available means to gauge user behaviour and thus, recommendation system using social networking will go a long way in helping the user to actually find out what they are looking for in a less cluttered manner. [2]

In Developing a Dynamic web recommendation System based on Incremental Data Mining Dr.V.V.Krishna et al. pro-posed a novel algorithm, called modified IncSpan for the effectual mining of the sequential patterns from the incremental database. The modified IncSpan algorithm can be able to discover the sequential patterns from the incremental database based on the sequential patterns obtained from the insert and append database and the closed sequential patterns are obtained from the resultant sequential patterns. The experimental results of our proposed web recommendation system showed that the proposed system outperformed with good precision, applicability and hit ratio. [1]

In An architecture of a Web recommender system using social network user profiles for e-commerce Damian Fijakowski et al. proposed a concept of a web e-commerce system that collects and uses, in the process of making recommendations, data obtained from social network profiles of its users. This ar-chitecture modeling approach was developed within the project of a mashup Web application that integrates with Facebook API. Their Further works should focus on defining weights for particular groups of keywords. Weights will represent significance of these groups and will have influence on values of keywords in those groups in recommendation process. [4] In Dynamic Recommendation System Using Web Usage Mining for E- commerece Users Prajyoti Lopes et al. focuses on providing real time dynamic recommendation to all the visitors of the website irrespective of been registered or unregistered. Action based rational recommendation technique is proposed that makes use of lexical patterns to generate item recommendation. It would be interesting to evaluate the proposed technique with different conventional recommendation approaches and measures its accuracy. This proposed system can also be tested for other application areas like movie recommendation, music recommendation etc. [5]

In Big Data Analysis: Recommendation System with Hadoop Framework Jai Prakash Verma et al. proposed a recommendation system for the large amount of data available on the web in the form of ratings, reviews, opinions, complaints, remarks, feedback, and comments about any item (product, event, individual and services) using Hadoop Framework. We have implemented Mahout Interfaces for analyzing the data provided by review and rating site for movies. Resultant graph is showing that whenever file size is increasing the execution time is not increasing in the same ratio and we know that data size that are in the form of ratings, ranks, review, feedback are increasing drastically. Here they are proposing Recommendation by applying the weightage of summarized reviews and opinions on the rating of item as future enhancement in this work. [6]

---

3. PROBLEM DESCRIPTION

Sometimes recommendation system hinders the scalability this means when large amount of data available at that time system ca not gives the scalable results. It may chance to upsetting their users because user can not get the desire item or product. System has difficult to filter the user interest related item from large data. So one problem is scalability, it also degrades the system throughput.
Another problem is time complexity. Sometimes system takes too much time for filtering process. So, it is also become the big problem because it takes much more time. Here we can also take accuracy parameter in consideration because recommendation system should gives the accurate product to its users which user exactly wants. So, we wish to work on one of above problem in my research area..

4. PROPOSED WORK

4.1 Proposed System Flow-Diagram

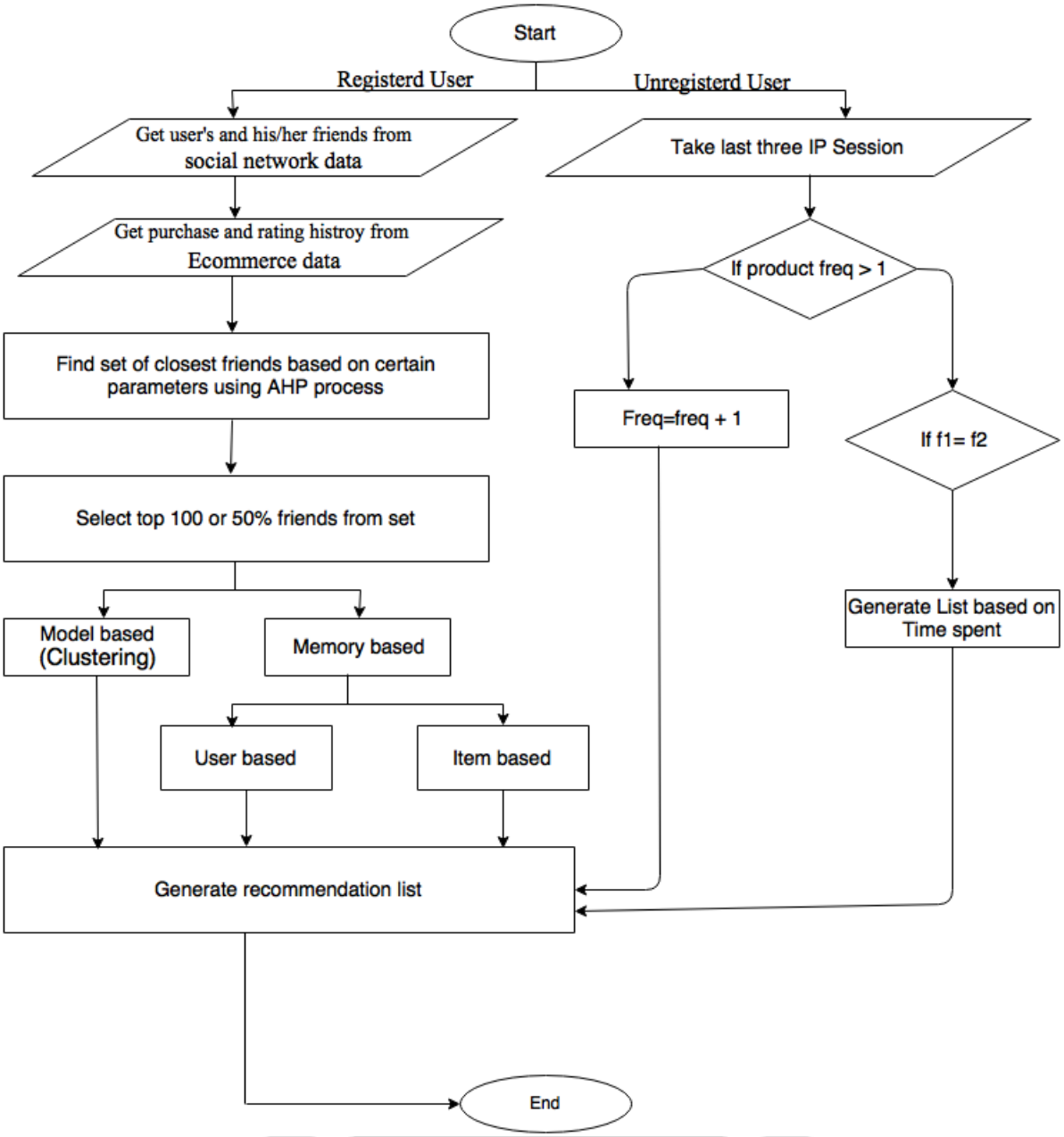

Figure 1: Proposed System Flow Diagram

Proposed solution steps describe as below:

1. First step is start the process and get user's and his/her friends social network data from websites.

2. After that in next step get purchase and rating history we can also say that e-commerce data from e-commerce websites.

3. Now, find set of closest friends based on certain parameters, like (age, language, gender) using Analytic hierarchy process- AHP.

4. Once we get closest friends then in next step select 100 or 50% from set.

5. Apply hybrid collaborative filtering approach (Memory based + Model based) for producing recommendation list to its users.

6. Finally system generates the recommendation list including anything like, products, items etc.

5. Experimental Result and Analysis

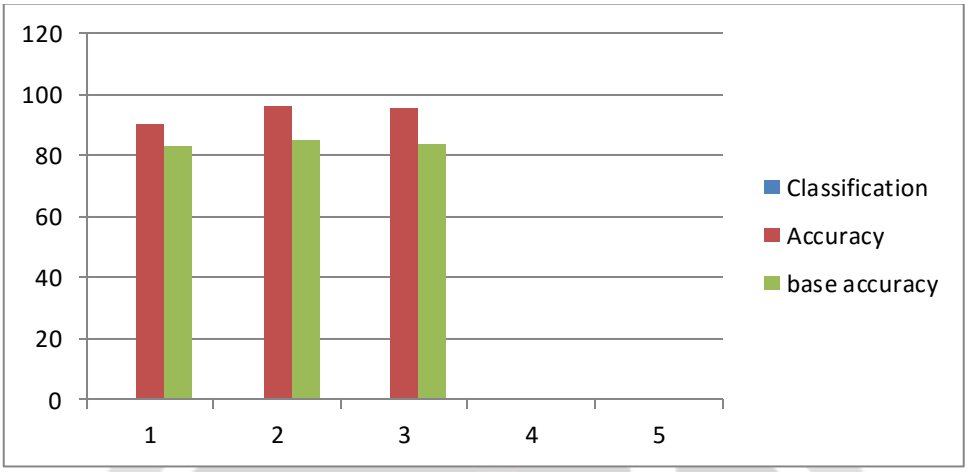

Chart 1 :-Accuracy comparison of base paper and proposed system

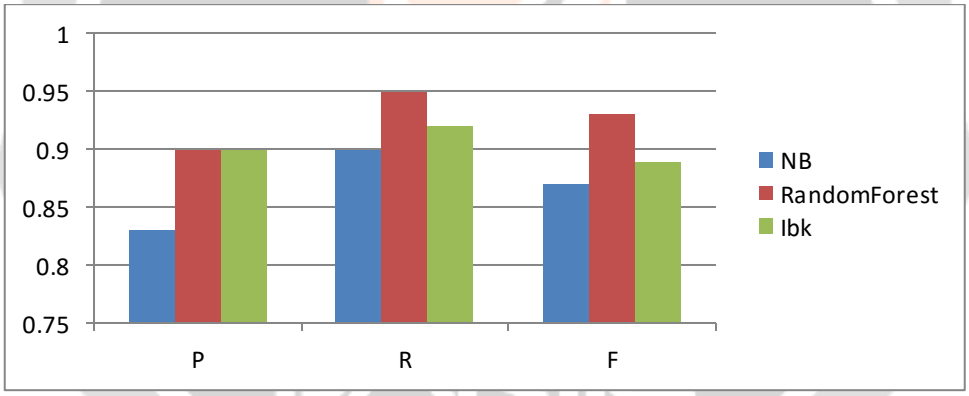

Chart 2 :-Precision, Recall, F1 measure for Proposed system

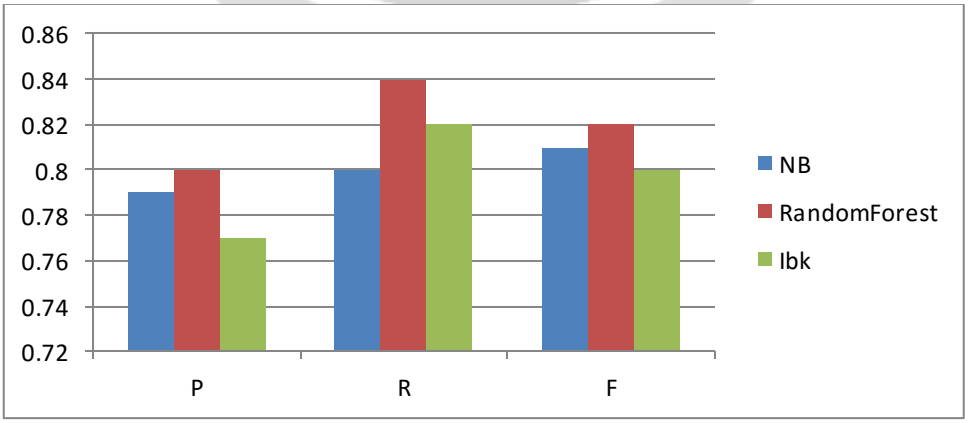

Chart 3 :- Precision, Recall, F1 measure for Base system

	Recommended Items	Not recommended items
Relevant	True positive(TP)	True negative(TN)
Irrelevant	False positive(FP)	False negative(FN)

Table 1. Recommendation Matrix[5

6. PARAMETERS:

$$Precision = \frac{True\_Positives(TP)}{True\_Positives(TP) + False\_Positives(FP)} \quad (1)$$

$$Recall = \frac{True\_Positives(TP)}{True\_Positives(TP) + False\_Negatives(FN)} \quad (2)$$

$$F1\ measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

These are the three parameters for measure accuracy result for Effective Recommendation.

7. CONCLUSION

One conclusion I can draw from this is that for making recommendation system more accurate and strong we have to use other information related to user from social networking websites like twitter, facebook. We can get better quality of results by using enhance techniques and algorithms instead of using traditional approaches. Hence I need to design proper mechanisms for Recommender systems from such challenges and risk such as scalability, time consuming process. So, this research work is concerned with study and analysis of recommendation techniques to extract the information from dataset and recommend appropriate product to its users.

REFERENCES

[1] Utpala Niranjan, Kanduri Srividya, Dr.V.V.Krishna, and Dr.V.Khanaa, Developing a Dynamic web recommendation System based on Incre-mental Data Mining, In Proceedings of IEEE, 2011.

[2] Md Zeeshan Ashraf, Dheeraj Kumar Chouwdhary, Rohan Lal Das, Prasun Ghosal, An Efficient and Optimized Recommendation System Using Social Network Knowledge Base, IEEE, 2014.

[3] Jyoti Pareek, Maitri Jhaveri, Abbas Kapasi, Malhar Trivedi, SNetRS: Social Networking in Recommendation System, Springer, 2013.

[4] Damian Fijakowski, Radosaw Zatoka, An architecture of a Web rec-ommender system using social network user profiles for e-commerce, IEEE, 2011.

[5] Prajyoti Lopes, Bidisha Roy, Dynamic Recommendation System Using Web Usage Mining for E-Commerce Users, IEEE 2015.

[6] Jai Prakash Verma, Bankim Patel, Atul Patel Big Data Analysis: Rec-ommendation System with Hadoop Framework 2 IEEE,2015 pp. 92-97

[7] Mukta kohar, Chhavi Rana Survey Paper on Recommendation System IJCSIT,2012

[8]Recommender system Wikipedia,'Recommendersystem',2015. [Online].Available:https://en.wikipedia.org/wiki/Recommender system# Collaborative filtering.[Accessed:06-Dec-2015]