Scientific Journal Impact Factor (SJIF) : 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

# Generic Recommendation Engine in Distributed Environment

Prof. Sujit Ahirrao[1], Faiz Akram[2], Swapnil Bagul[3], Kalpesh Modi[4], Harpreetkaur Saini[5]

[1,2,3,4,5] *Department of Computer Engineering*
[1,2,3,4,5] *Sandip Institute of Engineering & Management, Nashik(India).*

**Abstract**—Aim of this paper is to make recommendation engine very generic and allow user to run the recommendation engine on their data set in very small amount of time. For that Spark's scalable machine learning algorithm is used as core engine which will give the power to run machine learning algorithm in distributed environment and so run on large volume of data(Big Data).

**Keywords**—generic; distributed; spark; machine learning;

## I. INTRODUCTION

The huge increment in data available over the Internet is a challenge for searching of useful information, thus intelligent approaches are needed to provide users to easily locate and retrieve information from the internet. Now a days recommender engines, recommend everything from movies, books, music etc. Collaborative filtering (CF) algorithms are one of the most efficient recommendation techniques which suggest items and products as per user's interest. Collaborative filtering finds the relationships between the new user and the existing data in order to find the similarity and returns recommendations. There are two types of Collaborative filtering: 1.user-based Collaborative Filtering 2. item-based CF. User-based Collaborative Filtering finds a certain user's interests by finding other users who have likely interests and item based Collaborative Filtering checks a set of items rated by all users and finds how similar they are to the target item referred for recommendation.

[24]These techniques aim to fill within the missing entries of a user-item association matrix. Machine Learning library presently supports model-based cooperative filtering, within which users and merchandise are delineated by a little set of latent factors which will be wont to predict missing entries. Machine Learning library uses the alternating statistical procedure (ALS) rule to find out these latent factors. The implementation in Machine Learning library has the subsequent parameters.

**numBlocks** is that the variety of blocks wont to set computation (set to -1 to auto-configure).

**Rank** is that the variety of latent factors within the model.

**Iterations** is that the variety of iterations to run.

**Lambda** specifies the regularization parameter in ALS.

**ImplicitPrefs** specifies whether or not to use the express feedback ALS variant or one custom-made for implicit feedback knowledge.

Alpha could be a parameter applicable to the implicit feedback variant of ALS that governs the baseline confidence in preference observations.

## II. MACHINE LEARNINGCONFUSER-BASED

Machine learning develop a computer programs that can learn themselves about grow and change when exposed to new data. There are two types of machine learning.

 

### *A. Supervised Learning*

Supervised learning is the machine learning task of deducing a function from labeled training samples. In supervised learning, each example is a pair of an input object and a expected output value.

### *B. Unsupervised Learning*

Unsupervised Learning tries to find hidden structure in unlabeled data. Samples given to the learner are unlabeled. There is no error signal to evaluate a potential solution. This is the main difference supervised and unsupervised learning.

## III. LITERATURE SURVEY

Recommender systems have become an important research field since the emergence of the first paper on collaborative filtering in the mid-1990s. Although academic research on recommender systems has increased significantly over the past 10years, there are deficiencies in the comprehensive literature review and classification of that research.

**Bartosz Kupisz Et. Al[3],** was intended to develop and compare recommendation systems by using item-base collaborative filtering algorithm. The development was based on Hadoop and Spark. Data were collected from a social portal where users rate the items. Mahout library was used for Hadoop implementation. The solution was implemented using Apache Spark and Scala programming language. Tanimoto coefficient was considered as similarity measure because it provides the most precise results for the available data. As Apache Spark platform proved to be more efficient initial assumptions were concluded.

**Uma Sahu Et. Al[4],**As E-commerce Websites are increasing, it is very difficult for customers desired products and services. In order to solve the problem of data overload, personalized recommendation engines were developed to recommend desired product to customer with some relevant data. In this project a book recommendation system was developed which uses user or item based recommendation from Mahout. It will recommend books based on what similar users did in past.

**Zhiyang Jia Et. Al[5],** This work consist of recommender system which is constructed as online application and it generates a personalized list of attractions which user may like. The modern technologies like collaborative filtering are considered effective in tourist domain. The recommendation process of attractions divided into three steps: 1. Representation of tourist information. 2. Finding similar tourists. 3. Recommending attractions. The cosine method is used for finding similar tourists. And then recommendations are generated by referring history of similar tourists.

**Yingya Zhang[6],**Its waste of time to deal with huge recruiting information on internet. To deal with this problem a recommendation system for online-job hunting was developed. In this project a user-based and item-based collaborative filtering was contrasted. Students' resumes, recruiting details and weights of similar users were taken into consideration for designing recommendation system. The verification was done using experimental study using original data. It was found that recommended results can achieve high score of precision and recall, and having relevance with users' preferences in past.

**Anand Shanker Tewari Et. Al[7]**,Recommendation engines are helpful in finding appropriate products in e-commerce. Enormous amount of online book sellers lead to tremendous competition. This project presents a online book recommendation system for students. The characteristics of classification, association rule mining and user based collaborative filtering were combined in the system. The motive was to recommend books according to price range of user and publisher's name.

 

**Fudan Zheng[8],**As the network of internet getting more complex day-by-day number of websites having similar functions is also increasing. It will be better if users are provided with appropriate web services. This work presents a web recommendation system in accordance with collaborative filtering. The steps followed are: 1. Calculate users and service similarity using historical QoS raising information of websites. 2. Predict QoS values for current users in accordance with previous similar users' information. 3. Based on predictions do recommendations. For verification real world experiments were done.

**Yuncheng Li, Jiebo Luo Et. Al[9],**They worked on a genuine problem of image recommendation for web search engine users. Web search engine have became an integrated part of peoples' life. A two-stage method was developed to label users' choices for images by crowdsourcing technique. The stages are: 1. Finding out users general interests. 2. Predict whether user will be interested in image or not.Along with this they implement a baseline algorithm to establish the promise of the proposed cross-domain recommendation framework.

**Xiao Peng Et. Al[10],**The goal of this paper is to removing cons of traditional collaborative filtering algorithm. It proposes recommendation system based on a hybrid user and item-based collaborative filtering algorithm. When the system was applied on movieLens database the results were found to be improved.

**Robina Parveen Et. Al[11],**Recommendation system are useful to identify the subset of suitable learning resource from various choices. Recommendation system is research field since last few years. In this paper they define E-learning recommendation system. E-learning recommendation is part of recommendation system in which large amount of resources are available either offline or online. The aim of concept is to select the appropriate piece of material for study.

**Li Zhang Et. Al[12],**The collaborative filtering technique are useful for network recommendation system but the large amount of distributed data cause degradation of performance. The solution for this is to use the user based collaborative filtering algorithm which is used in this paper. This can be done using two ways:1. User similarity calculating method. 2. User item rating matrix extended. The result of this algorithm describe the user similarities more accurately and reduce the effect of data distribution to great extent.

**Yaming ZHANG Et. Al[13],**Collaborative filtering undergoes through sparse data problem and absence of scalability. To address these problems this paper proposed a distributed recommendation mechanism for mobile commerce. This system is based on cloud computing. Various experiments proved the efficiency of this system by overcoming scalability and sparse data problem.

**Badrul Sarwar Et. Al[14],**Recommendation system use KDD technique to solve problem of personalized recommendation. This paper examines various item-based recommendation algorithms. They studied various techniques for finding item-item similarities and different techniques for obtaining recommendation them. Finally, results were evaluated and compared them to basic k-nearest neighbor approach. It was observed that item-based algorithm works better than user-based algorithm.

***CHENG* Qiao Et. Al[15],**Current simulation resource management systems contain huge number of resources makes system lazy for use. This paper designs recommendation system with collaborative filtering which recommend appropriate simulation resource to user in relevance of his previous choices. The system has three components: 1. Collection of user past ratings. 2. Finding similar users 3. Recommend simulation resources. Pearson correlation is used to find similarity between various users. Then based on similar users' prediction of resources for current user is done.

**Rong Hu Et. Al[16],**Various cloud services are evolving in cloud market and it's a challenge to select appropriate one. This paper suggests use of recommendation engine with collaborative filtering. But in case of multi-functional system user may use some of the functions only and rate services. Such rating expresses user's preference for function instead of service. Thus it is advisable to measure similarity at granularity of functions' ratings. This paper used approach where user's rating is assigned to a function which user uses. After that recommendation is done by using a hybrid user-based and item-based collaborative filtering algorithm.

**Makoto Ichii Et.al[17],**Collaborative filtering algorithm is also use for Software component retrieve recommendation system .In this widely used software components are retrieve .In this paper recommendation system is incorporated with software retrieval system. For this they used browsing history to recommend appropriate components to users.

**Zheng Wan.al[18],**The Mobile commerce has become research topic in from last few years because of its commercial value & technical maturity. The most of service provider are participating in mobile commerce to increase its popularity & problem it's known as "spam recommendation" & decreases the users gratification .In this paper reduce the arbitrary recommendation system using recommendation filtering system. The aim is to record the user's operation i.e. "implicit feedback" on recommendation. In this the short message is obtaining user's interests & update user profiles endlessly. The instance when new recommendation are received then the system compute the resemblance between user profile & recommendation message & verify either the message is a spam or not.

**Zhi-Dan Zhao[19],**Collaborative filtering algorithm is use in most of the recommendation system, but the computational complexity of the collaborative filtering is high . So this obstruct use of recommender engine in large scale systems. In this paper they used the user based collaborative filtering on a cloud computing platform such as hadoop, for solving the scalability issue of collaborative filtering. The result of this simple method that partition users into groups on the two bases i.e proper arrangement of mapper number to overcome the initiation mapper & divide task evenly such that all processors end up task at similar instance, can accomplish linear acceleration.

## IV. PROPOSED SYSTEM

To make recommendation engine very generic and allow user to run the recommendation engine on their data set in very small amount of time. Use of Spark's scalable ML algorithm as our core engine which will give user a power to run machine learning algorithm in distributed environment and so run on large volume of data(BigData).

Let M be a recommendation system such that
R={D,N,M,C,O,L,U,I,R | ϕM}

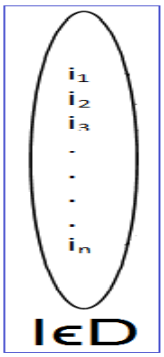


**figure 1**

where,
D is input database
D={i1,i2,i3.....in}
IϵD
i1=user_id
i2=product_id
i3=rating

N is normalized dataset
N={j1,j2,j3....jm} where m<=n

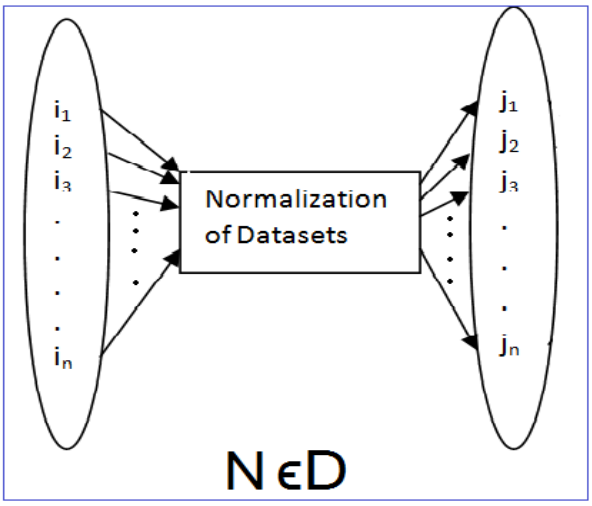


**figure 2**

M is user-item association matrix

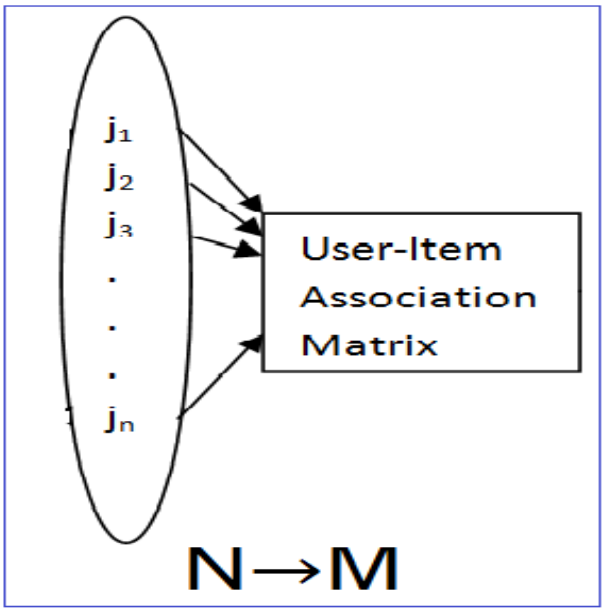


**figure 3**

C is collaborative filtering algorithm

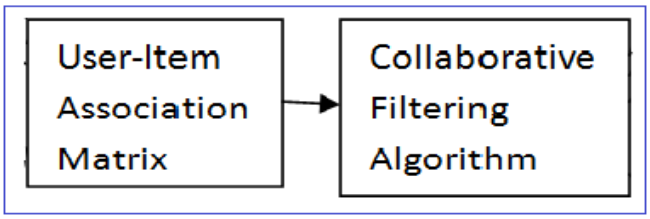


**figure 4**

O is set of missing attributes of use

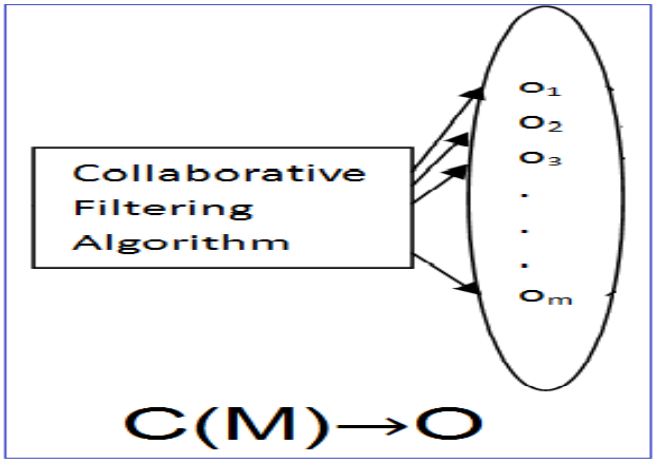


**figure 5**

L is machine learning model
U is set of user Ids
U={u1,u2,u3.....uk}
u1=1000
u2=1001
u3=1002
I is user information

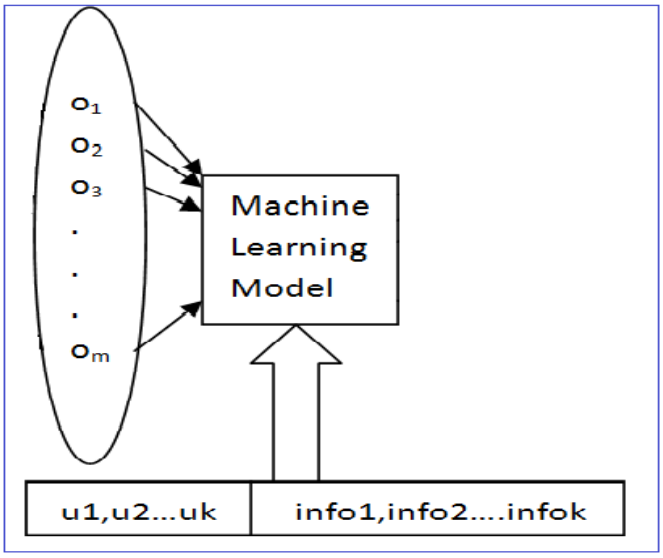


**figure 6**

R is set of ratings
R={r1,r2,r3.....rk}
r1=5.0
r2=2.0
r3=3.0
M(U,I)→R

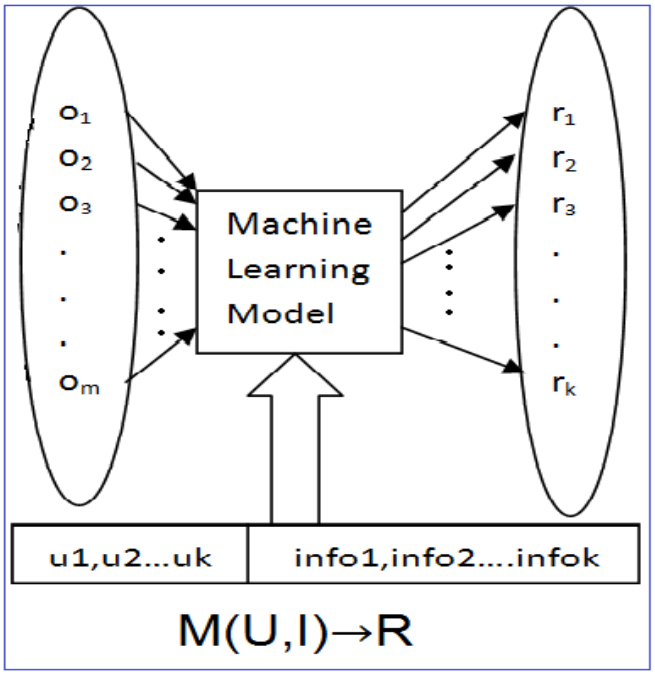


**figure 7**

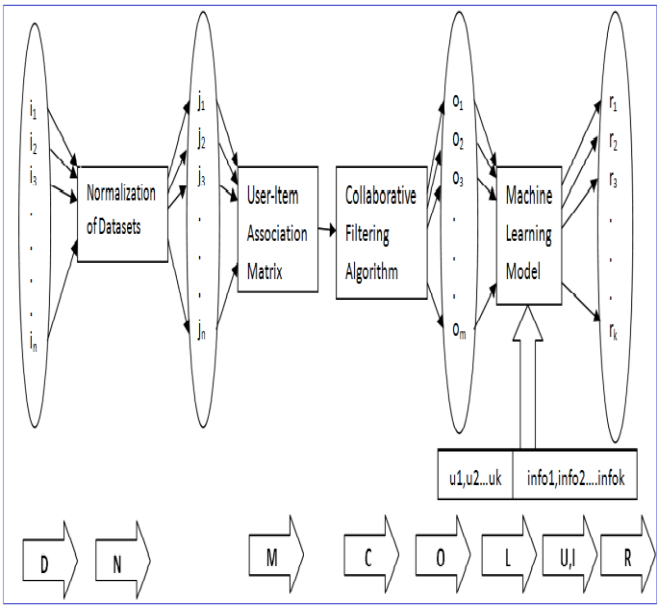


**Figure 8: Mathematical Model of Proposed System**

## V. FUTURE SCOPE

The system can be made more dynamic, in which the updates on profiles can be performed strictly in real time and also the innovations of hardware designs advance the computational speed, algorithms

 

and techniques with low time computational complexity are expected in the recommendation system developments.

It is difficult to keep the profiles up-to-date, online computations require many resources, such as memories and computational power. Therefore, it is important to maximize the offline computations.

## VI. CONCLUSION

Recommender systems have become ubiquitous. People use them to find books,music, news,movies. But all these recommender systems are application specific. In this paper we propose generic recommender system which can work recommend items for various databases that client provides.

## REFERENCES

[1] IEEE paper on A collaborative filtering recommendation engine in a distributed environment" by Ghuli, P.; Ghosh, A.; Shettar, R. Contemporary.

[2] IEEE paper on Hybrid Recommendation System based on Collaborative Filtering and Fuzzy Numbers" by Miguel A. G. Pinto, Ricardo Tanscheit,Marley Vellasco Department of Electrical Engineering Pontical Catholic University of Rio de Janeiro Rio de Janeiro - Brazil 2012.

[3] Bartosz Kupisz Et. Al "Collaborative Filtering Recommendation Algorithm based on Hadoop and Spark" ,published by IEEE 2015,pp 1510-1514 .

[4] Uma Sahu Et. Al 1. "Personalized Recommendation Engine Using HADOOP",International Conference on Technologies for Sustainable Development (ICTSD-2015), Feb. 04-06, 2015, Mumbai, India,pp 1-6 .

[5] Zhiyang Jia Et. al "User-based Collaborative Filtering for Tourist Attraction Recommendations ",IEEE International Conference on Computational Intelligence & Communication Technology (2015),pp 22-25.

[6] Yingya Zhang "A Research of Job Recommendation System Based on CollaborativeFiltering",Seventh International Symposium on Computational Intelligence and Design(2015),pp 533-538.

[7] Anand Shanker Tewari Et. al "1. Book Recommendation System Based on Collaborative Filtering and Association Rule Mining for College Students",published by IEEE 978-1-4799-6629-5/14/$31.00_c 2014,pp 135-138.

[8] Fudan Zheng "A web service recommendation approach based on Collaborative filtering ",IEEE 11th International Conference on e-Business Engineering(2014),pp 345-349.

[9] Yuncheng Li , Jiebo Luo Et. Al " PERSONALIZED IMAGE RECOMMENDATION FOR WEB SEARCH ENGINE USERS " Published by IEEE(2014),pp 1-6.

[10] Xiao Peng Et. al "Improved Collaborative Filtering Algorithm In The Research And Application Of Personalized Movie Recommendations",Fourth International Conference on Intelligent Systems Design and Engineering Applications92013),pp 349-352.

[11] Robina Parveen Et. al "E-Learning recommendation system –A survey",International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN : 2278-800X, www.ijerd.com Volume 4, Issue 12 (November 2012), PP. 10-12.

[12] Li Zhang Et. al "An improved collaborative filtering algorithm based on user interest ",JOURNAL OF SOFTWARE, VOL. 9, NO. 4, APRIL 2014 ,pp 999-1007.

[13] Yaming ZHANG Et. al "A Distributed Collaborative Filtering Recommendation Mechanism for Mobile Commerce Based on Cloud Computing ",Journal of Information & Computational Science 8: 16 (2011) ,pp 3883-3891.

[14] Badrul Sarwar Et. al "Item Based Collaborative Filtering Recommendation Algorithms ",published by IEEE (2013),pp -11.

[15] *CHENG* Qiao Et. Al "Simulation Resource Recommendation System Based on Collaborative Filtering ",published by IEEE 978-1-4799-0333-7/13 © 2013,pp 448-452 .

[16] Rong Hu Et. al "A Hybrid Collaborative Filtering Approach for Multi-functional Service Recommendation",IEEE Third International Conference on Cloud and Green Computing (203),pp 54-59.

[17] Makoto Ichii Et.al "Software Component Recommendation Using Collaborative Filtering",: *SUITE'09*, May 16, 2009, Vancouver, Canada 978-1-4244-3740-5/09/$25.00 □□2009 IEEE,pp 17-21.

[18] Zheng Wan.al "Design of a Recommendation Filtering System in Mobile Commerce",published by IEEE(2009),pp 1-4.

[19] Zhi-Dan Zhao "User-based Collaborative-Filtering Recommendation Algorithms on Hadoop ",Third International Conference on Knowledge Discovery and Data Mining (2010),pp 478-481.

[20] Industry Report on"Amazon.com Recommendations Item-to-Item Collaborative Filtering" by Greg Linden, Brent Smith, and Jeremy York.

[21] Lang K. (1995). News Weeder:Learning to filter net news.In Proceeding of the 12th international conference on machine learning, Tahoe City ,CA.

[22] Dhoha A., Ghadeer S., Lamia A., Mona K., Romy M.A Survey Paper on recommender Systems.www.arxiv.org/computer science /Information retrival,2010.11.

[23] Karypis,G.(2000).Evaluation of Item- Based Top- N Recommendation

[24] Algorithms. Technical Report CS- TR- 00- 46,Computer Science Dept.,University of Minnesota..

[25] https://spark.apache.org/docs/latest/mllib-collaborative-filtering.html

 