A Product Recommendation System Using HADOOP Server

Ms. Neha Sehta[1], Abhishek Patidar[2]
H.O.D. of I.T.[1], M.E. Student[2]
Department of CSE
Sushila Devi Bansal College of Technology, Indore, India

Abstract:
Bigdata analysis is the approach for analyzing extremely large dataset with minimum computation effort and resource requirements. It applies to revel the trends, pattern, association especially relating to human behavior and item characteristics. The E-commerce is one of the widely used techniques in day-to-day life of many individuals. Each websites are trying to business their economy more and more. However, these sites are with wide varieties of complexities. The data on such shopping sites is very large that when it will try to manage and manipulate it, it becomes very tougher. A Product Recommendation can help to enhance the business. Here, study has been performed to evaluate the need of product recommendation and use of it using Hadoop server. This paper has considered the importance product recommendation and limitations in existing system.

Keywords: Big data, Collaborative filtering, E-commerce, Hadoop, Product Recommendation.

I. INTRODUCTION

Big data can be defined as collection of heterogeneous data from variety of sources. It can be defined as the data from large organizations like business, web. Thus, managing thus large database is tough task with the help of traditional tools. The major issue with such large data is that it includes integration, visualization, storage, searching. Data analytics on such large data is quite tough. The analysis of data is demand of the technology as popularity can be calculated using it. Today the hidden pattern discovery is major demand of research. As the variety of resources is available, therefore the need is to merge them all to utilize it in better way[1].

The E-commerce is one of the widely used techniques in day-to-day life of many individuals. Each websites are trying to business their economy more and more. However, these sites are with wide varieties of complexities. The data on such shopping sites is very large that when it will try to manage and manipulate it, it becomes very tougher. Wide variety of product is available in these shopping sites. E-commerce sites are gaining attention this day in order to get the product delivery just at home. The product available in such sites varies from small pin to large vehicle. Anything, which one can imagine of, can be available in such sites. There should some techniques to process such large data. The data available in these sites are arranged in the form of ratings, ranks, feedback, remarks etc. The aim of such sites is that more economy to be earned with wide variety of availability. In large business websites, the data is stored in big data form; if the results are analyzed in the form of big data then patterns drawn will result in useful knowledge extraction[2].

This sites also use the web portals in order to get the opinion from the user about their services and based on the response occurred the result are generated. The recommender system uses this for having even better recommendation. The growth of e commerce sites is mainly depends upon the recommendation system. This recommendation system uses the ratings system and uses the summary of feedback for more productivity. The big data is used for storing database; hence, the need is to have analysis on such ecommerce sites to have better recommendation[3].

The figure represents the architecture of big data. As we know that big data can manage large quantity of data to efficiently, handle such massive volume of data. Apache Hadoop is open source software, which can manage large data warehouse in multiple nodes of clusters. The Map Reduce programming model is used in Hadoop, the scalable and distributed architecture can be easily managed in it. The beauty of Hadoop is that it provide volume, velocity, variety, veracity.
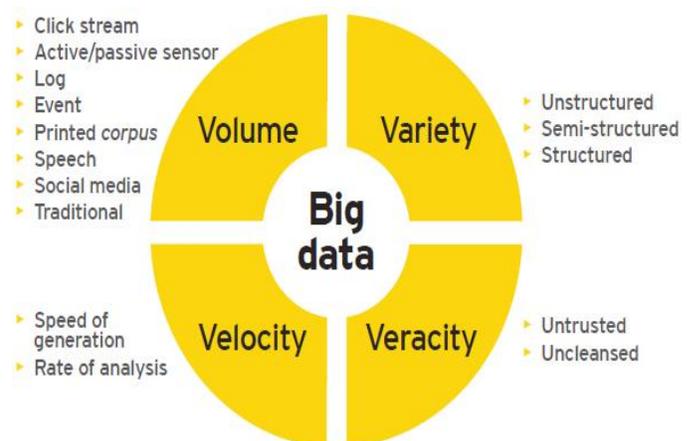

Figure 1: 4-V's of Hadoop [1]

In our research, we are focusing on product-based recommendation. There are many products available in shopping sites; the aim is to have proper recommendation with efficient approach. We are designing the Hadoop based cluster in which the algorithm for clustering is designed names along with the making of custom classifier. In our approach big data analysis using Hadoop is performed, the filtering performed is collaborative filtering. Hadoop uses map reduce algorithm, the concept of map reduce is that initially mapping of the needed items is performed then only relevant one will be reduce. The database used in Hadoop is HDFS, which is commonly called as Hadoop distributed file system. In it single node algorithm and multiple node or distributed algorithm are used in processing of data. The big data processing can be done using such algorithm. In product-based analytics, the optimization can be performed using HBase, which is nothing but the NoSQL database. The need of HBase is to provide better performance and low latency analytics.

II. PROMINENT RESEARCH WORK

Anindita A Khade [1] proposed a solution in which product based recommendation is provided. The given research mainly works on Hbase database. The main component are Hadoop nodes and distributed recommendation engine. The dataset used is of amazon. Distributed recommender is merged in it to have better applications. Product data is stored in big data format. The working of key component can be given as under:
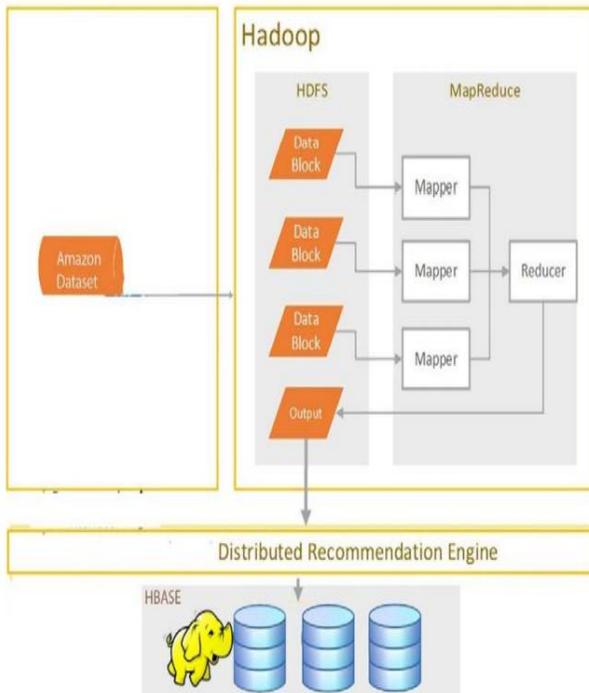

Figure 2: Hadoop Architecture

When the data from mining is analyzed the review in it are used for making the recommendation better. The review consist of rating, rating determine which review is highly favored by users. Every customer having certain id therefore it is find that which customer send that, the Amazon work on the Hadoop cluster in Map reduce format. The customer are having suggestion based on the data which they prefer,

hence it is never necessary that two customer have recommendations. The scenario work in map reduces form hence key and value related to it are generated in HDFS. Hence, the customary classification is form in order to have the variable input to different customer.
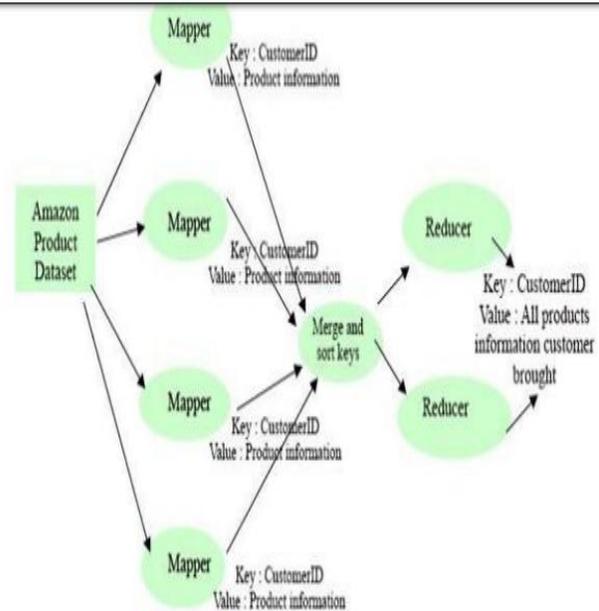

Figure 3: Map-Reduce Architecture

Data formatter is used in order to convert it in key value format, this key value pair is given to map function. The key is this reference is customer id whereas the value id the product that is referred is value in the pair. Here comes the new entity called as reducer, the work of this reducer is to collect the entire key and generate the value pair with respect to it. The reducer always has the customer profile, which is responsible for having the entire product, which is purchased by any user. Hence, the recommendation provided is based on the customer interest. HDFS is Hadoop distributed file system in which the server is kept at various sites, so if any system crashes then other server will start working the process. In addition, the process consist of big data hence query is distributed to sites in order to process it in bulk. In this the aim is to used Hadoop in big data of Amazon shopping sites in order to find their pattern as well as in order to give them best recommendation on the basis of their interest. The language used in our work is python, which is one of the very well known language of machine learning. For large dataset map reduce is used because of its efficiency along with less overhead. In our work the collaborative filtering is used as algorithm and data is stored on Hadoop cluster for analysis.

Author [2] observe that customer analysis is performed where customer of shopping sites is considered. The big analytics is performed, as he data from such customer is very large. There are many ways of classifying the customer the classification can be based on decision tress where it is also possible that classification can be based on clustering. Customer is analyzed based on various parameters like sales, inventory planning, fraud detection, market optimization etc. In this approach the decision trees are used. In order to enhance the performance and handling of various types of

data classification is performed. For better visualization this decisions tree are designed and root nodes and leaf nodes have certain classification hierarchy in that. The decision tree of same type of data is formed. The best part of decision tree is that it provides visualization. Hence, customer-buying patterns are observed.

The flow of the system is as follows:

1) HDFS consist of customer details, which are given as an input in desired algorithm.
2) Decision tree algorithm are used hence it is designed and invoke an instance.
3) MapReduce is then performed in order to get the current node and its instance is same or not. For the entire attribute, which are not covered in Map Reduce, are used in class label of instance and output the index.
4) Reduce function measure the index and value of class label combination and count is returned.
5) Parameters associated with like entropy, information gain and gain ratio of attributes is then calculated.
6) The input is then processed using Map reduce approach.
7) Certain decision rules are decided then and storing of it in HDFS is performed.
8) Using user interface the new data is input.
9) The decision rule will then category of that data and placed it in decision tree
10) The dataset will then have the visualization of it is observed using decision tree.
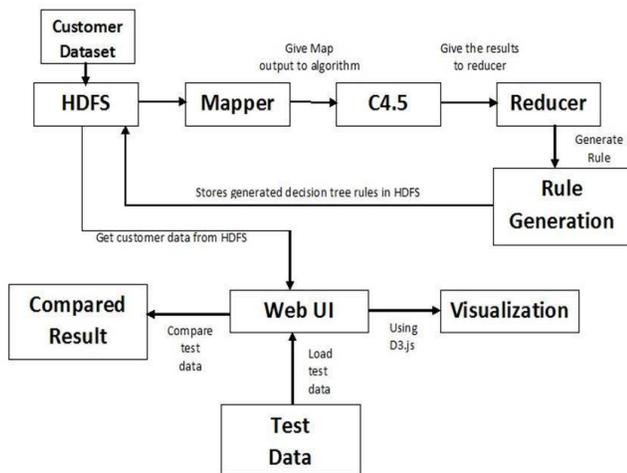11) Calculate entropy, information gain and gain ratio of attributes.

Figure 4: Existing System

The distributed architecture is used in this approach and decision tree along with MapReduce is performed. The data visualization is then performed, the map reduce is used keeping in mind that the big data is present, and analytics on big data cannot be opt using easy approaches. Cloud based architecture is used in order to access it from anywhere.

Author [3] Praveena Mathew uses the concept of book recommendation system in which content base filtering and collaborative filtering is performed. The mining is performed using the concept of association rule, hence a hybrid approach is used in order to recommend the relevant books. The effective algorithm is the result.

Certain steps are involved in the algorithm:

Step 1: The book dataset is collected and the process of data integration is performed. We first perform the process of scanning, book dataset is then cleaning, and unnecessary data is removed by data cleaning.

Step 2: Data Preprocessing
The data, which is desired for processing, is only used. Hence in this process categories and subcategories are formed.

Step 3: Filtering Transactions
Filtration is performed like we are demanding only books of specific university then they are filtered.

Step 4: Content based Filtering
Perform Classification based on different measures like classes; subject or any other is performed. Ratings of books are associated in it.

Step 5: Perform Collaborative Filtering
Each resource is having some quality. The quality of book based on content is performed.

Step 6: Final Recommendations

Based on user choice and quality, the recommendation is given. The book of user interest is then revealed.
The book recommendation system work same as product recommendation in which the recommendation is given based on user choice of book. The classification of book is performed initially then classification is custom based classification having the concept of association rule and collaborative filtering also. The history of user is checked and recommendations are given. At the end the recommendation is given using the book suggestion on the basis of user profile as well as quality of book. The concept can be useful in academic and research purposes.

Author [4] JunBo Xia concentrates on the issue of product based recommendation, in its Ecommerce based recommendation is provided and issue regarding product recommendation is solved. The algorithm used are collaborative filtering as well as custom classification. The user preference are consider and best possible recommendations are given. Hence, desired product is recommended when user performs any search.

The recommendation results of Ecommerce product recommendation are calculated as follows.

$$\phi_{u,x,p}^{a,v} = \phi_u^{a,v} + \gamma \cdot \sum_{i=1}^{m} Sim(u_i, u_j) \cdot (\phi_{u_j,x,p} - \phi_{u_j}^{a,v})$$

Where p and u mean the product recommended results and users. $v \_ g(x,a)$;

III. COMPARATIVE STUDY

PAPER TITLE	STRENGTH	WEAKNESS	APPLICATIONS
Performing Customer Behavior Analysis using Big Data Analytics [1]	Provide facility to recommend based on customer behavior. This solution can be useful for web personalization	1. Do not consider the product view option. 2. It does not provide any web personalization option for dynamic product view	Dynamic Product View and Web Personalization.
A Scalable Product Recommendations using Collaborative Filtering in Hadoop for Bigdata[2]	Collaborative Filtering is finest method for recommendation purpose. It provide help to find similarity among product and can be useful for prediction of sales and upcoming targets	Performance estimation has been performed for single node cluster. It does not consider any customer view or nature behind purchasing	Classification of similar product and category. Can be used for Low cost recommendation.
E-commerce Product Recommendation Method based on Collaborative Filtering Technology[3]	Collaborative filtering solution for e-commerce portals. Evaluate the perfromce for MovieLen dataset.	Only consider small dataset. Do not involve any distribution policy.	Useful product for E-commerce portals for product recommendation.
Book Recommendation System through Content Based and Collaborative Filtering Method[4]	An hybrid algorithm to integrate content mining and collaborative filtering for book recommendation. Provide opportunity to find similar nature books	Content mining can only useful for descriptive mining. It can't be consider as for product recommendation or analysis purpose	Very useful for descriptive document based recommendation
Combining Big Data Analytics with Business Process using Reengineering	Provide approach to integrate bigdata analytical approach with reverse engineering.	Can't be feasible for recommendation purposed	Understanding of big data processing and handling.
Performance Enhancement of Hadoop MapReduce Framework for Analyzing BigData[5]	Processing of Large data with distribution policy	Do not provide any specific solution for recommendation purpose	It can be useful for tuning of job configuration

IV. PROBLEM DOMAIN

Our study from previous work analyze that there are lot of online shopping portals use in day-to-day life of an individual. With increase in demand of at home shopping, every large organization is trying to vast their business. It is realized that online shopping sites does not provide shopping related features it also perform through analysis on customer profile. Hence when the discount are given based on the nature of product. Once the widely used product and user nature is analyzed then the recommender comes into frame. The portal start giving recommendation on the basis of user nature. Either in current days system the recommendation is used based on one perception the merchant is focused or the customer. The recommendation is provided on the basis on history of user profile. However, merchant profit are never considered.

V. PROPOSED SOLUTION

Our solutions provide the way to have recommendations in way that is more appropriate. The recommendation is more approach is provided using all the factors in mind.

In our work, initially the classification is performed based on decision tree. This classification of decision tree is custom classification in which certain specific characteristics is considered. The collaborative filtering is then performed, the aim of using it once the classification is completed, and the further suggestion should be such that it should be desired quality. For this rating and ranking is done and highly rated result is used.

The aim of our work is that when the customer search for any product, then all parameters like user history, related product is recommended. When recommender provides any result it is observed that what kind of product that customer usually referred like price based or quality based. After that based on custom classification the collaborative filtering in performed after it. The concept is to have on the first classification; the remaining entries should have top ranked result as collaborative filtering gives the quality-based result by using the concept of ranking.

The entire work is performed using Hadoop cluster in which Map Reduce approach is used. The Map reduce framework uses the concept of key value pair in order to have reduced result. The entire scenario is implemented in Amazon site dataset. The ultimate result is recommendation of each user in more way that is refined.

VI. CONCLUSION

Our work concludes that the recommendation is very much necessary for user as well as merchant point of view. Our scheme provides desired recommendation in efficient way. As decision tree is used, the visualization can be realized. Hence, our system is better than the designed architecture and customer satisfaction is achieved by suitable recommendation. The computation overhead can be reduced in future work.

REFERENCES

[1] Anindita A Khade, "Performing Customer Behavior Analysis using Big Data Analytics" published in 7th International Conference on Communication, Computing and Virtualization 2016.

[2] Riyaz P A, Surekha Mariam Varghese, "A Scalable Product Recommendations using Collaborative Filtering in Hadoop for Bigdata" published in International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015).

[3] Fatima Rodrigues, Bruno Ferreira, "Product recommendation based on shared customer's behaviour" published in Conference on Enterprise Information System / International Conference on Project Management / Conference on Health and Social Care Information System and Technologies, CENTERIS / projMAN / HCist 2016, October 5-7, 2016.

[4] Ms. Praveena Mathew1," Book Recommendation System through Content Based and Collaborative Filtering Method" .

[5] JunBo Xia, "E-commerce Product Recommendation Method based on Collaborative Filtering Technology" published in International Conference on Smart Grid and Electrical Automation 2016.

[6] Meena Jha, Sanjay Jha, Liam O'Brien, "Combining Big Data Analytics with Business Process using Reengineering".

[7] JunBo Xia, "E-commerce Product Recommendation Method based on Collaborative Filtering Technology" published in International Conference on Smart Grid and Electrical Automation 2016.

[8] Amazon Dataset: http://jmcauley.ucsd.edu/data/amazon/

[9] Mohammed Alodib, Zaki Malik "*A Big Data approach to enhance the integration of Access Control Policies for Web Services*" IEEE ICIS 2015, June 28-July 1 2015, Las Vegas, USA.

[10] Xianfeng Yang and Liming Lian,et. al., 2014, "A New Data Mining Algorithm based on MapReduce and Hadoop," Xinxiang University, Xinxiang Henan, P.R.CHINA.