Survey on Recommendation System in Big Data Analytics

Dr. Golda Dilip[#1], T.P.Ezhilarasi[*2], M.Easwari[#3], M.Alice Janovier Gayen[*4]

[#1]*Associate Professor,* [*2]*Assistant Professor,* [#3,*4] *UG Scholar*

Department of Computer Science and Engineering

[#1]*Saveetha School of Engineering(SSE), Saveetha University,Cheenai.*

[*2,#3,*4]*RAAK College of Engineering and Technology,Puducherry.*

Abstract-**Recommendation system is employed to predict the user interest supported the user style and suggest things to the user supported the likes and reviews of various things. the big quantity of increase in data obtainable over the web has created a greatest challenges in looking helpful data. There area unit several knowledge filtering algorithmic program area unit obtainable for recommendation system like user primarily based filtering , item primarily based filtering and hybrid filtering that create that are attempting to predict the user interest with greatest accuracy. however there exist some problems in handling streaming knowledge. therefore we have a tendency to area unit process a cooperative filtering recommendation system with Hadoop and Spark framework to processed recommendation to the user with higher accuracy.**

Keywords-Recommendation system, hybrid filtering, Hadoop, Spark framework.

I.   INTRODUCTION

We are living in 100% data world. There are various data generating factors are available such as sensors, cctv cameras, social network, online shopping, airlines, hospitality data ect. These are the various data generating factors. This way of getting huge amount of data. This huge amount of data can be store in a local server the processing speed is not good. This data is called big data. Now a days data will be increased rapidly in last two decades due to increase the number of internet user. In 2013 more than 350,000 tweets per minute, more than 400 hours of video uploaded for each minute, facebook users click the like button on more than 4 million posts every minute, around 4 million google search conducted by every minute. 205 billion emails are sent each day in 2015, in approximately 2020 will be produced by 44 zettabytes of data. The big data analytics is the process of examining large

and varied data sets i.e. big data. In processing the large amount of data the efficiency and scalability is less to avoid this problem by using hadoop and mapreduce technology. Fig 1.1 shows the use of big data analytics.

Fig 1.1 use of big data analytics

In 1990, google was started by how the way of storing large amount of data in efficiently. After 13 years google was introduced by GPS in 2003. In 2004 introduced in best processing technique such as mapreduce. Mapreduce is used to processing the data which is stored in HDFS, and which is used for analyzing and retrieving big datasets efficiently. By processing the large datasets by using large number of nodes. The collection of node is called cluster. If all the nodes can be used in same local network. Mapreduce contains to main parts such as job tracker and task tracker. Task tracker is attached to a data node, the data node managing the replication of data in three node. Job tracker is act as a manager it will control all the node. The mapreducer and be divided into a three parts such as, map(), shuffle() and reduce().Map function is used to map all the same word and shuffle is used to reordering the collection

of words and reduce is used to group of processed data with key. Fig 1.2 shows the mapreduce architecture.
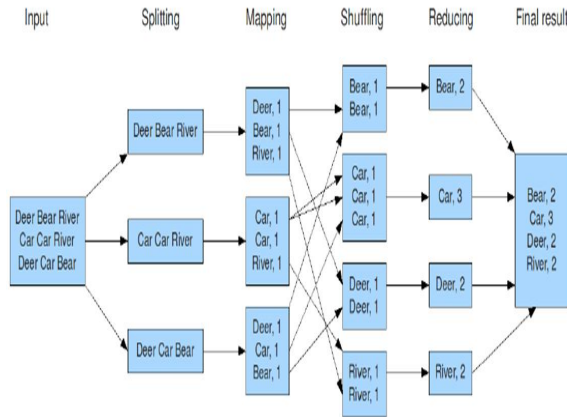

Fig 1.2 Architecture of mapreduce

In 2006-07 yahoo was introduced by HDFS. Hadoop distributed file system is used to store the large amount of data. Mapreduce is used to processing the data in HDFS. The HDFS and Mapreduce is the core concept of Hadoop. Hadoop is a framework and it is a open source. It is used to provide the more efficiency and scalability of data. Hadoop is an Java-based programming framework that supports the processing and storage of extremely large data sets in a distributed computing environment. Hadoop is the best solution for big data, it provide the proper storage of local system. Which is provided by more efficiency and scalability it is store and process huge data in less time. It can be manages the thousands of terabytes of data. It is a distributed file system which is replicate the data in various node if the node is failure the replicated data can be used. It can be handle various set of data such a business data, sensor data, scientific data ect.

Recommendation system could be a data filtering technique, it's wont to offer the data he/she could also be interested. It is provided by the relevant item supported user interest. the information are collected from user history. that product is often likable or purchased supported that information to advocate the merchandise. the advice system relies on information assortment, rating and filtering. These 3 area unit chiefly used for recommendation system. information assortment is that assortment of user information that relies on user sort of a product, buying the merchandise and rating the merchandise. Rating is very important a part of recommendation system, as a result of it sense what a user feel regarding the merchandise. It reflects the likes of, adding to a

cart, buying or simply clicking. the utmost rating is five. Filtering is that the product supported rating and alternative user information, the filtering is classed into 3 varieties like

Collaborative Filtering: It can be used to comparing of users. If two user purchase the same product then if the user A purchase any product will be recommended to user B.

Content Based Filtering: It can be item based recommendation that are similar to liked in the product past, it can be represented by a set of descriptor or terms.

Hybrid Filtering: It is combination of user based and item based filtering. To combine multiple recommendation techniques to produce the output Fig 1.3 shows the user and item based recommendation.
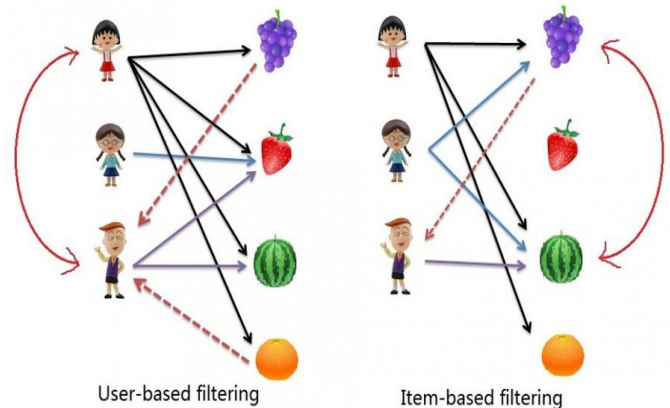

Fig 1.3 User based and Item based Filtering

II. LITERATURE SURVEY

Service Recommendation Techniques:

Vijay M Bande, Ganesh K Pakel,[1] proposed a technology to utilize service recommendation techniques, which goes for showing a customized service recommendation set and recommender ssystems are recommending the foremost valuable services to end-users. To improve efficiency and scalability issue in massive data surroundings, it is implemented in hadoop framework which uses map reduce parallel processing model. The disadvantage of this paper is does not consider user similarities.

Dynamic Recommendation Technique:

Prajyoti Lopes, Bidisha roy, [2] proposed a real time dynamic recommendation technique to all the visitor of the website irrespective of been registered or unregistered user. In this paper proposed Action based rational recommendation technique that make use of lexical patterns to generate item recommendation. In this paper provided by effectiveness of a collecting real time Ecommerce data and comparing the system with user based and product based technique. The disadvantage of this paper is produce more noise data and occurring on the false positive error.

Hotel Recommendation System:

Khushboo Ramesh Shrote, Anil V. Deorankar, [3] Proposed a hotel recommendation system based on hadoop framework. The score calculation can be applied for sentiment analysis. The calculated score will be used for recommendation purpose. The hadoop framework is mainly used to increase the efficiency and scalability of a recommendation system. The main disadvantage of this paper is cold start problem can be introduced.

Hybrid Recommendation Technique:

Arupananda Girish Prasad,Mahendra Kumar Gourisaria,Lalit Kumar Vashishtha, [4] proposed a paper to used MapReduce framework and comparing the differences time consumption of common seral hybrid recommendation algorithm with parallel hybrid recommendation algorithm by using different clusters. Here we have applied the experiments by using Each Movie data sets to exploit the advantages of parallel algorithm. Also from the experiments we can get how our improved parallel K-Mean algorithm by using two methods such as distance measure method and initial centroids method which are based on MapReduce can achieve higher accuracy as compare to tradition K-Mean algorithm and also our performance can improve. The disadvantage of this paper is less efficiency and scalability.

Micro-Video Recommendation System:

Piyush Gupta, Atul Sharma, Jitender Grover, [5] proposed a paper to a micro-video recommendation system. The recommendation algorithms are the core of this system. Traditional recommendation algorithms include content-based recommendation, collaboration recommendation algorithms, and so on. At the Bid Data times, the challenges what we meet are data scale, performance of computing, and other aspects. Thus, this paper improves the raditional recommendation

algorithms, using the popular parallel computing framework to process the Big Data. Slope one recommendation algorithm is a parallel computing algorithm based on MapReduce and Hadoop framework which is a high performance parallel computing platform. The other aspect of this system is data visualization. Only an intuitive, accurate visualization interface, the viewers and producers can find what they need through the micro-video recommendation system. The disadvantage of this paper is evaluated using small data sets.

Rating Based Mechanism:

Songtao Shang, Minyong Shi, Wenqian Shang, Zhiguo Hong, [6] proposed a paper a rating based mechanism that distinguishes abnormal posts with the help of users rating. If rating is positive then post is normal otherwise it is abnormal. To implement proposed mechanism we used hadoop platform and MapReduce paradigm. The disadvantage of this paper is only user rating are considered.

III. CONCLUSION

We have studied in detail various issues of recommendation systems. We have analyzed the existing systems and found that they are not scalable enough. Hence, using hadoop with spark framework and collaborative filtering can ensure good scalability and processing the streaming data. The spark frame work is used to provide the best performance, ease of use, cost effective and more security of data.

Reference:

[1] Vijay M Bande, Ganesh K Pakel, CRSR:Customized service recommendation system for bigdata analysis using mapreduce, IEEE(2017).
[2] Prajyoti Lopes, Bidisha roy, Dynamic recommendation system using web usage mining for E-commerce users, Science Direct(2015).
[3] Khushboo Ramesh Shrote, Anil V. Deorankar, Hotel Recommendation System using Hadoop and MapReduce for Big Data , IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol.6, No.2, Mar-April 2016
[4] Arupananda Girish Prasad,Mahendra Kumar Gourisaria,Lalit Kumar Vashishtha, Building Hybrid Recommendation System Based on Hadoop Framework, International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016
[5] Harshit Patel1,Pooja Jardosh, Intelligent Recommendation System Using Big Data and Hadoop, IJARIIE(2016).
[6] Piyush Gupta, Atul Sharma, Jitender Grover, Rating based Mechanism to Contrast Abnormal Posts on Movies Reviews using MapReduce Paradigm.
[7] Songtao Shang, Minyong Shi, Wenqian Shang, Zhiguo Hong, A Micro-video Recommendation System Based on Big Data,IEEE(2016).