Graph based Recommendation for Distributed Systems

Vivek Pandey
Dept. of Computer Science and Engineering
Shri Shankaracharya Technical Campus
Junwani, Bhilai(C.G.)

Padma Bonde, PhD
Dept. of Computer Science and Engineering
Shri Shankaracharya Technical Campus
Junwani, Bhilai(C.G.)

ABSTRACT
A huge amount of information available through electronically, the requirement for effective information retrieval and the implementation of filtering tools have become more necessary for easy retrieval of relevant information. Recommendation Systems (RS) are the software tools and methods providing recommendation for items as well as services to be of need to a user. These systems are providing widespread success in e-commerce applications now-a-days, with the generation of internet. This paper presents a survey of the area of recommendation systems and illustrates the state of the art of the recommendation technique that are generally classified into three categories: Content based Collaborative, Demographic and Hybrid systems. This paper discusses the advantages and disadvantages of the current survey categories as well as the trustworthiness of the recommendation system in a new dimension as searching the evaluator for more suitable recommendations. In the domain of recommendation system, this work can also help to put forward for the use of researcher as an enabling technology.

Keywords
Recommendation System, Hadoop, MapReduce, Collaborative, Graph Based.

1. INTRODUCTION
Recommendation systems is one of the subclass of information filtering system that helps to make prediction on 'rating' or 'preference' that a user would help to provide an item. The concept of recommendation systems was come into existence to deal with the problems of information overload, to scan through the huge information datasets, and helps to retrieve the most relevant information [1]. For example, while recommending any product, service and location that can be book, video, music, research resources, TV programs, documents, and website. With the help of personalized recommendation.Recommendation system helps to recommend about any items or services to the user on the basis of their purchases of similar products or services by the other customers. Gathered information creates some potential problem like information filteration of technique is needed i.e. Recommendation system. Information need to be prioritized for the user rather than just filtering the right information, to retrieve these we require to classify data accurately [2].

There are variety of classification methods for classifying both multivariate and univariate datasets. Some of the classification methods are, Decision tree classifier, Bayesian classifiers, Support vector machine (SVM), K-Nearest-Neighbor classifier (KNN).

To handle huge amount of information (also called Big Data), we need a robust system. Now the question arises: How do we examine this vital amount of data? The most appropriate solution to this is with Distributed framework. Hadoop is an open source framework which helps to develop and executes distributed applications which can process huge amounts of data. Big data is a term that illustrates large volumes of high velocity, complex and varied data that need advanced methods to enable the capture, storage, distribution, management, and analysis of the information [3].

The 3 Vs. of Big Data management:

- Volume: There are huge data than ever before, it's continuously increasing size.
- Variety: There are various different types of data available such as text, sensor data, audio, video, graph, etc.
- Velocity: Data is in continuous as streams of data, and we are interested in determining useful information from it in real time [4].

Hadoop has its own file system known as HDFS (Hadoop Distributed File System). HDFS utilizes master and slave architecture. HDFS logically distinguishes the file system metadata as well as application data. The metadata is stored on a delicate computer known as NameNode (known as master node) in HDFS [2].

Application data were stored on other computers named as DataNodes (known as slave node). A data file is categories into one or more blocks and these category blocks are replicated and stored across various DataNodes. All of the nodes present in a Hadoop cluster. Which are fully connected with each other. The NameNode consists of the file system namespace and the mapping of file blocks to DataNodes.

When an HDFS client reads a file, it first contacts the NameNode to obtain the locations of data blocks comprising the file and then access these data blocks from closest DataNode. Applying the concept of map reduces where mapper class work as practitioner and reducer class work as a combiner [5]

2. LITERATURE SURVEY
Introduced [6] a new approach to improve the quality of collaborative filtering recommendation systems. The algorithm combines item clustering and weighted slope one scheme. We compare our approach to other algorithms on the Movielens dataset. The results suggest our algorithm produces better result. In the future, we should to introduce the other automatic method to determine the two parameters of DBSCAN. And we also can use larger data sets to evaluate the availability of the algorithm.

[7] Introduced two semantic social recommendation algorithms called Node-Edge-Based and Node-Based, these algorithms recommend an input item to a group of users. In fact, we assume that, users are connected via collaboration social network, and users and items are described via semantic taxonomy. Our proposed algorithms are mainly

based on Depth First Search algorithm with some modifications that are related to firstly, the semantic similarity between users and the input item. And secondly, the social network analysis measures. We applied these two algorithms on a real dataset (Amazon Dataset), and we compared them with the collaborative filtering and hybrid algorithms. Our results showed that Node-Edge-Based gives the best F-measure values, with a good performance.

The paper [8] used a combined approach of user-user CF and item-item CF has been presented to generate recommendations on Hadoop cluster using Apache Mahout, a library for machine learning algorithms. By using combined approach, accuracy of recommendation has improved. This approach has scaled well with the hadoop platform. Time needed to solve the problem has reduced. Mahout is able to handle big data but it still lacks some algorithms. The recommendation for single user need to be improved for better results. New computing platforms like Apache Spark are getting prominent in the field of Big Data analysis. Recommendation algorithms can be performed on such platforms for faster performance.

[9] Proposed a Recommendation engines are a natural fit for analytics platforms. They involve processing large amounts of data, using the machine learning algorithms. Now days, the data size is increasing with the unstructured format, it is not possible to handle with the Mahout. It combine Apache consumer data that collected online, and the results of the analysis feed real-time online applications. Hadoop is being increasingly used for building out the recommendation platforms. This paper mainly focuses on evaluating the classification accuracy metrics using the Apache Hadoop and Mahout. By experiments it conclude that, Mahout can handle large amount of structured Hadoop and Mahout for the recommendation, it can recommend large amount of structured and the unstructured data efficiently and fastly.

Paper [10] presented on combined Collaborative Filtering using Mahout on Hadoop for movie recommendation. By combining User-based and Item-based Collaborative filtering, accuracy of the results gets improve. Hadoop has increased throughput. Because of multiple computer nodes, time taken for solving problem has been reduced. Mahout is feasible for handling large amount of structured data. But now a day's data are more unstructured. Hadoop using Mahout can handle big data statistically. To handle real time data randomly, HBASE, HIVE are the better solutions.

3. VARIOUS EXISTING MECHANISM AND THEIR LIMITATION

The Clustering algorithm [6] proposed a recommended framework where it joins item clustering and weighted slope [13] one plan. It utilized Movie Lens Data set to demonstrate the outcome. The strength of this algorithm is it can improve the accuracy of the collaborative filtering recommendation system. In any case, the primary disadvantage of the calculation is just used little Data set.

The Node–Based algorithms [7] and [10] used amazon data set with three main features: a social network analysis measure (degree centrality), the graph searching algorithm (Depth First Search algorithm), and the semantic similarity measure (which measures the closeness between an input item and users).It is observed t that Node-Edge-Based gives the best F-measure values, with a good performance.

Collaborative Algorithm [8] of the combined approach of user-user collaborative filtering and item-item introduced to create suggestions on Hadoop bunch utilizing Apache Mahout, a library for machine learning calculations. The performance of the algorithm is worstcause it used sparse matrix.

Paper [9] concentrated on contrasting the different comparability estimation calculations and order exactness measurements on Hadoop and non-Hadoop condition used Apache Mahout and item-based collaborative filtering.it additionally situated in light of cooperative calculation with Movie Lens Dataset. In any case, the impediment of the calculation is nonexclusive calculation which is not adequate. Collaborative Filtering utilized Mahout on Hadoop for Movie proposal. By consolidating User-based and Item-based CF, the precision of the outcomes gets improved. The main quality of this strategy is time lessens to fathom the extensive scale problem. But not proficient while managing real-time datasets[10].

4. CONCLUSION

Data in the form of feedback, remarks, reviews, opinions, and complaint treated as huge amount of data which is generally known as Big Data which cannot be used directly for recommendation system. These data first needs to filter/transform as per requirement. This is survey of papers and problems related for handling data. These paper survey different concepts of recommendation system on distributed Hadoop framework.

5. REFERENCES

[1] Yang Xiaoshan; Zhu Ligu; Zhang Qicong and Feng Dongyu,"A Reserch on Big Data Storage Utilization", 24th International Confrence on Applied Computing and Information Technolog, 2016 Las Vegas, NY, USA, pp.368-372.

[2] Satya Ranjan Dash, Satchidananda Dehuri2, "Comparative Study of Different Classification Techniques for Post-Operative Patient Dataset", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 1, Issue 5, July 2013.

[3] B.Sarwar, G.Karypis and J. Konstan, "Item-based collaborative filtering recommendation algorithms" , Proceedings of the 10th international conference on World Wide Web,pp.285-295.

[4] Fatos Xhafa , Adriana Bogza and Santi Caballé, "Performance Evaluationof Mahout Clustering Algori-them Using a Twitter Streaming Dataset "IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), 27-29 March 2017,Taiwan

[5] Kunhui Lin, Jingjin Wang, Meihong Wang, "A Hybrid Recommendation Algorithm Based on Hadoop", the 9th International Conference on Computer Science Education (ICCSE 2014) 22-24 August , Vancouver, Canada.

[6] Haipeng You, Hui Li, Yunmin Wang, and Qingzhuang Zhao, "An Improved Collaborative Filtering Recommendation Algorithm Combining Item Clustering and Slope One Scheme", Proceedings of the International MultiConference of Engineers and Computer Scientists Vol 1,pp 313-316, 2015 Honkong.

[7] Dalia Sulieman, Maria Malek and Dominique Laurent, "Graph Searching Algorithms For Semantic-Social

Recommendation", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012

[8] Yongchang Wang and Ligu Zhu , "Reserch on Collaborative Filtering Recomndantion Algorithem Based On Mahout", Proceedings of The 4th Intl. conf.on Applied computing and Information Technology, 2016.

[9] Dr. Senthil Kumar Thangavel, Neetha Susan Thampi, Johnpaul C I, " Performance Analysis of Various Recommendation Algorithms Using Apache Hadoop and Mahout", International Journal of Scientific & Engineering Research, Volume 4, Issue 12, December-2013

[10] G.R.Bamnote and Agrawal,"Evaluting and Implementing collaborative Filtering Sysystems Using Appache Mahout",International Conference on Computing Communication Control and Automation, 2015.

[11] Hongqing Guo,Li Chen A personalized recommendation technology for E-commerce website" Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, 10-13 July, 2016,Jeju, South Korea,

[12] M.Bell, and Y.Koren , "Scalable collaborative filtering with jointly derived neighborhood interpolation weights" ,Data Mining, Seventh IEEE International Conference , Aug 2015pp.43-52.

[13] Mamadou Diaby, Emmanuel Viennet and Tristan Launay "Toward the next generation of recruitment tools: An oonline social network-based job recommender system" Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on 25-28 Aug. 2013 Niagara Falls, Canada.

[14] Jingjin Wang, Kunhui Lin,and Jia Li, "An Collaborative Filtering Recommendation Algorithm Based On User Clustering and Slope One Scheme", Proceedings of The 8th International Conference on Computer Science Education(ICCSE) Colombo Srilanka.