

A Website Recommender System using Hadoop for Particular Domain

B. S. V. Swaroop¹, Amit D. More²,
Siddhesh B. Kadam³, Shubham W. Ingle⁴ and Prof. M. R. Patil⁵
Department of Computer Engineering
SKNCOE, University of Pune, Pune, India

Abstract: Hadoop is an open source framework for writing and running distributed applications that processes large amount of data. "A Website Recommender System using Hadoop" is implemented for recommending websites in particular domain. Website Recommender system mainly uses HDFS (Hadoop Distributed File System) and Map-Reduce function. Website assessment at every time becomes tedious for new user, thus to deal with this issue Recommendation method is present. So in this paper we discuss different data filtering techniques, interfaces that will be helpful to implement Recommender system.

Keywords: Hadoop, Recommender System, Map-Reduce, HDFS.

I INTRODUCTION

Hadoop is the most widely known and most widely used implementation of Map Reduce paradigm. Hadoop is a framework of tools for large scale computation and data processing of large data sets. A cluster of Hadoop consists of several machines and services. At least one node of Hadoop cluster should have HDFS service and Map Reduce service. In HDFS, a name node is a service which handles the task management, data assignment, and scheduling. Usually, the secondary name node is also established in case the primary name node fails to work properly. In the same manner, each node is able to take over other node when a failure occurs. For security reason, the data copied onto HDFS will be duplicated to multiple data nodes to increase their liability. This replication process also allows the ability of retrieving data from the nearest node. Website Recommender system uses HDFS for storing and distributing the data. There will be one master node and all others are slaves. Master node will handle all slave nodes. The unique user id will be assigned to the browser. Browser's history will be taken as input and it will be stored in database. Map-Reduce paradigm is used for filtering and aggregating the data. Map function will keep the count that access by most users and also with recent timestamp. Reduce function will aggregate the count and will give final result. [1]

Website Recommender System uses Map-Reduce results for recommending websites. To implement recommendation some criteria is required for recommending websites. So, Website Recommender System takes help of Multi-Criteria Rating system. Where this technique uses similarity based and function based approach. In the function based approach, 3-dimensional

multi-rating criteria is decomposed into 3-means single dimensional recommendation problems where rating can be estimated by applying Reduce function for aggregating known ratings and then we calculate the predicted rating. [3]

II LITERATURE SURVEY

A Map Reduce-Based Parallel Clustering algorithm for Large Protein-Protein Interaction Networks: Li Liu, Dangping Fan, Ming Liu, Guandong Xu, Shiping Chen, Yuan Zhou, Xiwei Chen, Qianru Wang, and Yufeng Wei proposed a map reduce algorithm for the clustering. The algorithm had basically 2 steps of map-reduce, first was the forward map-reduce and then backward map reduce. In this as they have taken the data set as the protein show it becomes like a graph. In the forward map-reduce they took some of the aspects to compute like colour of the node to determine it. This method starts with taking one of node as root and then processing it and checking the shortest path, it is done with each of the node and BFS is applied and distance is calculated. Then the backward map reduce algorithm is applied and which is done in parallel with each given an edge to calculate edge betweenness. Then with both the results of forward and backward the clusters are prepared.

Hybrid Parallel Approach for Personalized Literature Recommendation System: Kun Ma, Tingting Lu, Ajith Abraham proposed a method to recommend of published papers. In this they have taken aspects like researcher id, researcher time, information about browser. In this they first took the published paper data with the help of the crawlers and RSS listener. Then they categorized them using Latent Dirichlet Allocation (LDA). Then for further filtering they applied collaborative filtering technique for understanding user behavior. At last they finally applied matrix factorization with alternate least square and recommended each user the top N similar papers.

New Recommendation Techniques for Multi-Criteria Rating Systems: Gediminas Adomavicius, Young Ok Kwon proposed two new techniques for recommendation for multi-criteria rating those are similarity based and function-based approach. In similarity approach to extending standard collaborative filtering techniques they have taken three steps calculating distance between two users rating for same item then calculating the overall distance and applying cosine similarity formula. In

function based approach three steps are taken. It starts with decomposing k-dimensional multi-criteria rating space into k- means single dimensional recommendation problems and applying algorithm to estimate rating for each user then machine learning technique to estimate aggregation function of known ratings and then we calculate the predicted rating

Recommender Systems in E-Commerce: J. Ben Schafer, Joseph Konstan, John Riedl have compared between the different e-commerce sites of how they used to get information and recommend them. This paper basically focuses on three different aspects those are recommendation interface, technology and finding recommendations. In interface they have used E-mail, browsing, similar items, Average Rating Text Comments. In technology they used item to item correlation, attribute based, person to person correlation. In finding recommendation they have Organic Navigation, Keywords/freeform, Selection options, Request List.

Implementations of Web-based Recommender Systems Using Hybrid Methods: JanuszSobecki proposed the set of steps that can be used for web-based recommendation. In this paper they proposed two different consensus-based hybrid methods. Consensus theory has its general origins in the social sciences and in the theory of choice in particular. The first uses the mixture of demographic and collaborative filtering, in this the user is registered then he is made of the cluster by his profile and recommendation happens. In second there is additional joining of content based along with the demographic and collaborative making it more precise then compared to the individual using of different aspects.

III CONCLUSION

This paper focuses on different published papers which are required to build the website recommender system on Hadoop. It shows us different algorithms methods which can be used to implement the website recommender system more efficient. The concept of map reduce would be very helpful in the filtering of data and many concepts for filtering, interface and technology are mentioned in this paper.

REFERENCES

- (1) Li Liu, Dangping Fan, Ming Liu, Guandong Xu, Shiping Chen, Yuan Zhou, Xiwei Chen, Qianru Wang, and Yufeng Wei, "A Map Reduce-Based Parallel Clustering Algorithm for Large Protein-Protein Interaction Networks" 2012
- (2) Kun Ma, Tingting Lu, Ajith Abraham. "Hybrid Parallel Approach for Personalized Literature Recommendation System" 2012 Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan, China.
- (3) Gediminas Adomavicius, Young Ok Kwon, "New Recommendation Techniques for Multi-Criteria Rating Systems" Department of Information and Decision Sciences Carlson School of Management University of Minnesota.
- (4) J. Ben Schafer, Joseph Konstan, John Riedl, "Recommender Systems in E-Commerce" 1999 GroupLens Research Project Department of Computer Science and Engineering University of Minnesota Minneapolis, MN 554551-612-625-4002.
- (5) Janusz Sobecki, "Implementations of Web-based Recommender Systems Using Hybrid Methods" 2006 Institute of Applied Informatics Wroclaw University of Technology, 50-370 Wroclaw, ul. Wyb. Wyspianskiego 27, Poland.

BIOGRAPHY

- (1) B. S. V. Swaroop is student of SKNCOE, Pune, University of Pune pursuing B.E. Computer Engineering degree.
- (2) Amit Moreis student of SKNCOE, Pune, University of Pune pursuing B.E. Computer Engineering degree.
- (3) Siddhesh Kadamis student of SKNCOE, Pune, University of Pune pursuing B.E. Computer Engineering degree.
- (4) Shubham Ingleis student of SKNCOE, Pune, University of Pune pursuing B.E. Computer Engineering degree.
- (5) M. R. Patilis professor at SKNCOE, Pune, University of Pune.