A Implementation on Big Data Analysis for Mobile Marketing using Hadoop

#1Prof. Nikita Bhattacharjee, #2Prathamesh Kalyankar, #3Ishwar Wankhede,
#4Shubham Agarwal, #5Shridhar Jadhav

1nikitabhattacharjee@dypic.in
2pratham.kalyankar0@gmail.com
3ishwar.wankhade@gmail.com
4agarwalshubham610@gmail.com
5jadhavshridhar999@gmail.com

#12345Department of Computer Engineering,
Dr. D.Y.Patil School of Engineering

ABSTRACT

Most widely used frameworks for developing MapReduce-based applications is Apache Hadoop. But developers find number of challenges in the Hadoop framework, which causes problem for the management of the resources in the Map Reduce cluster that will optimize the performance of MapReduce applications running on it. There is Enormous growth in mobile marketing so most of the companies use the data mining. Daily a lot of data is generated which makes a number of available jobs related to big data marketing and advertising. This needs an up to date updation to clients and application dataset. In this paper we have studied the data mining techniques with the mobile marketing dataset based on Hadoop platform. We have analyzed advertising data of mobile users using Hadoop and recommend the next advertisement.

Keywords: Big Data, Data mining, Big data Analytics, Decision Tree, Hadoop

ARTICLE INFO

I. INTRODUCTION

Big Data is a collection of large and unstructured form of data. There are numerous characteristics of big data 1)Volume, 2) Variety, 3) Velocity, 4) Veracity. We study mobile marketing analysis whose dataset is collected from mobile marketing warehouse. This dataset consists of a large amount of mobile data such as category, name price, flow of records which reflect the details. Each record in the dataset consists of several user information such as advertisements, past internet history, to recommend what kind of advertisement has to be shown to user. By examining all the flow of data, we try to figure out and recommend what advertisement suits the best according to user's behavioural pattern. The apps or websites appearing together seem to be highly related, but there has not been found any explicit understanding of the association rules.

The popularity of mobile apps is rising dramatically due to the prosperity of smartphone market. As a tool to access the Internet, smartphones and tablets contribute a considerable scale of network traffic in daily life. According to the report of 2014 Internet Trends released by Kleiner Perkins Caufield & Byers (KPCB), the proportion of Asia is the highest and the Internet traffic on global scale has taken up to 25 percent so the volume of data will also increase, Liu Yan, et.al., [1] This papers intends to discover interesting patterns in this growing internet traffic. The goal is to identify strong association from frequently used apps and websites. In order to find association between apps and websites Apriori Algorithm is used.

The increase in use of mobile wallets has tremendous potential of finding behavioural pattern of user. It helps us understand what kind of products are purchased by particular user, Celina Alexandra, et.al.,[2]. This paper show CRISP-DM (Cross-Industry standard process for data mining) methodology which is used to obtain information by the purchases made by Mobile Wallet users. The mobile wallet ecosystem is intended for the dematerialization of cards which helps in improvement of advertising and marketing.

II. LITERATURE SURVEY

[1] James Manyika et al., "Disruptive technologies: Advances that will transform life, business and the global economy", This gives alternative techniques to entrepreneurs with technologies or methods to a market delivering. No one change the product trading ways. Source to a many number of different terms to improve and transform life trade and world economy. A drawback is, its

needs in-depth analysis of key potential is difficult. Try to adopt technologies may take years of time for different transforming methods.

[2] Giamas Alex, "Spark, Storm and Real Time Analytics", These are also aims like hadoop for the computation of distributed events. Distributed talks capabilities of computation are to be done with spark and storm as like hadoop. The advantages are its helps to computation of distributed computation capabilities. It is applied for large scale of data for processing written in scale presenting using Apache spark.

[3] Martin Kihn and Mike McGuire, "Gartner Webinars, Mobile Marketing and Data-Driven Marketing", As like a hadoop, spark is also an open source cluster calculating environment. Optimized interactive workloads for queries of a spark enabled in memory distributed datasets. The advantages are, unlike hadoop, integration between distributed data and manipulation is done with scale. Side by side over the hadoop file system for complementary of jobs distribution that support datasets by a third party clustering framework.

[4] X Lieu, "Knowing of Big Data Processing and analytics", First of all knowing of the processing and analytics of big data is important. In this we learned that how actually big data processing and analysis is done.

[5] Celina Alexandre, João Santos, Petru Bocsanean, David Monteiro, João Mangana, Patrick Santos, "Marketing Behaviors Analysis in a Mobile Wallet Solution Using Data Mining", in this paper author presented a data mining study for a novel mobile wallet ecosystem, where the behaviors and preferences of the consumers are analyzed in order to create targeted marketing campaigns through their mobile wallets. The CRISP-DM methodology was used to perform the data mining studies.

[6] Ying Li, Pavankumar Murali, Nan Shao and Anshul Sheopuri, "Applying Data Mining Techniques to DirectMarketing: Challenges and Solutions", In this paper, author discuss one of the most common tasks faced by marketers when faced with resource and time constraints, namely, consumer prioritization with the objective of optimizing one or more marketing key performance indicators such as consumer conversion.

[7] Purushottam R Patil, Pravin Revankar , Prashant Joshi, "The Application of Data Mining For Direct Marketing", In this Paper author proposes an application implemented in Java to integrate data mining and Campaign management Software. Decision tree Algorithm C4.5 to build Data mining model by taking inputs from a database to predict customer behaviour such as Loyal and Unloyal.

[8] Tássio de O. S. Auad, Luiz Felipe C. Mendes, "Improving the User Experience on Mobile Apps Through Data Mining", With the advance of mobile applications market, there is an increasing concern about the challenges when developing products that meet the many types of users and to harmonize each product with the various usage environments, thus, providing a good user experience.

[9] Weng Suxiang, Jin Yonsheng, "Deeply Analysis in Mobile Clients' Consumptive Behavior Based on Data Mining Technology", In this paper, author discuss and analyze the consumptive behavior based on data mining technology. Firstly, we fully analyze the characteristic of mobile clients' consumptive behaviors. According to these characters, not only one data mining method can be used.

[10] Liu Yan, Yu Ke, Wu Xiaofei, "Association Analysis Based On Mobile Traffic Flow for Correlation Mining Of Mobile Apps", This dataset is acquired by a traffic monitor which is deployed in the backbone network of a mobile network operator in a province of China. In this paper author analyze both the mobile applications and the websites which the mobile user has accessed. Then an association analysis method is conducted to extract rules which can reflect characteristics and preference of users when accessing the Internet by smartphones.

III. HADOOP WORKFLOW

A. Hadoop Workflow The workflow of Hadoop is shown in Fig. 1. It has the following steps:

1. Input text files to a platform.
2. Server portioning file to blocks with the same size, then assigns a block of data to each computing node. 3) The compute node runs map on the input data and producing intermediate data pair for every word, then sends its intermediate data pairs to the node designated to perform the reduce operation.
3. The reduce operation counts the number of occurrences of each word using the values and emits it as a key-value pair.
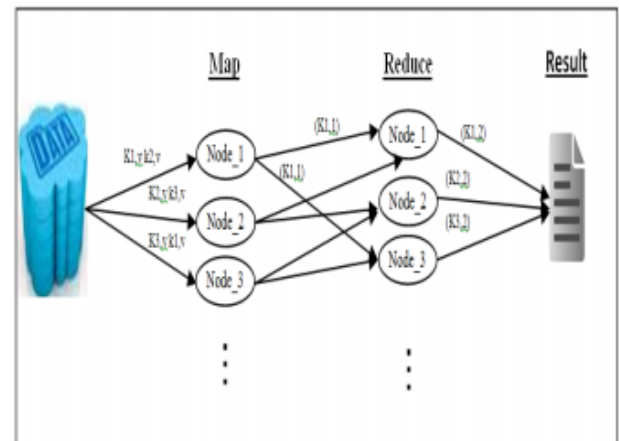4. Server receives the results and outputs the list.

Fig 1. Hadoop workflow

B. Hadoop Drawbacks

Hadoop Drawbacks From the architecture of Hadoop and its workflow of data computation, there are many drawbacks of Hadoop. These drawbacks are:

1) Hadoop needs high memory and big storage to apply replication technique.
2) Hadoop supports allocation of tasks only and do not have strategy to support scheduling of tasks.
3) Still single master (NameNode) which requires care

4) Load time is long. These drawbacks effect on both the performance and reliability of Hadoop against big data analysis. Therefore, it is necessary to develop a new framework or modify some Hadoop features to overcome Hadoop limitations and improve its performance and reliability. So, in this paper, a new framework is proposed to overcome the drawbacks of Hadoop and improve big data analysis.

IV. SYSTEM REQUIREMENTS

Hardware Requirements:-
- 3.2GHz Intel i3 processor
- 2 GB RAM

Software Requirements:-
- Apache Hadoop
- MYSQL Server
- XAMMP
- Eclipse IDE

V. ALGORITHM

Algorithm:
Input: Customer's Mobile Reviews

Output: Class of the Reviews/ Classified Reviews Processing:

Step 1: User Comment
    Obtain comments / reviews provided by the user for processing

Naive bayes algorithm
Step 2: Conversion of Review(Preprocess)
    i. Determine the user comment based on mobile product.
    ii. Generate dictionary for sentiment words using the clustered
    iii. Data clustered is done
    iv. Verify keywords

Step 3: Classifier()
    For training inconsistency classifier proceeds for pattern maching

Step 4: Pattern analysis()
    Then identifies word with two contexts words i.e positive, negative

Step 5: Prediction()
    The class of the reviews is specified from the set pattern.

Step 6: Finla analysis result()

Step 7: Stop

VI. RESULT

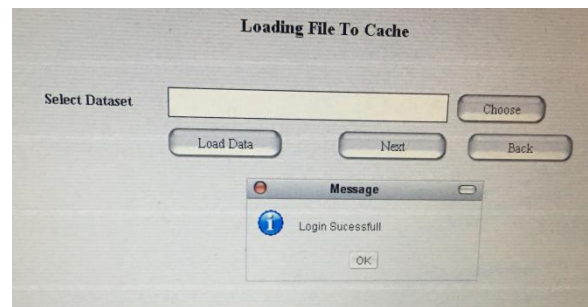Fig 1. Here load the mobile marketing data

Fig 2. Shows all mobile data for processing

VII. CONCLUSION AND FUTURE SCOPE

In this paper we investigated the techniques for analysing behavioural patterns of multiple users. Based on their usage of apps, internet and messages we recommend users the advertisements that best suits there interest.

For the future work, we will try to introduce more methods of analysis and measurement to inspect different dataset. Meanwhile, comparison should be introduced to perfect the evaluation and the association rules detection and evaluation can be potentially improved.

VIII. ACKNOWLEDGMENT

REFERENCES

1] Liu Yan, et.al., "Association Analysis based on mobile traffic flow for Correlation Mining of mobile apps", 2014

2] Celina Alexandra, et.al., "Marketing Behaviors Analysis in a Mobile Wallet Solution Using Data Mining", 2017

3] Chandrasekar Vuppalapati, et.al., "Building a Big Data Analytics Service Framework for Mobile Advertising and Marketing", Feb 2015.

4] Shaowei Lin, et.al., "Mobile Big Data Analytics Using Deep Learning and Apache Spark", May 30th,2016

5] Ying He, et.al., "Big Data Analytics in Mobile Cellular Networks", Feb 25th,2016

6] Prakash V Parande, et.al., "Big Data Analytics Service Framework for Advertising and Marketing",8th August 2016

7] Ravi Vatrapu, et.al., "Social Set Analysis: A Set Theoretical Approach
to Big Data Analytics", 29th March ,2010

8] Ying Li, et.al., "Applying Data Mining Techniques to Direct Marketing: Challenges and Solutions", 2015

9] Huaying Shu, et.al., "Integration of data mining techniques to evaluate promotion for mobile customers' data traffic in data plan", 22nd Sept 2016

10] Weng Suxiang, et.al., "Deeply Analysis in Mobile Clients' Consumptive Behaviour Based on Data Mining Technology", 10th Nov 2009