

A Survey on Big Data Analysis for Mobile Advertising and Marketing

Prof. Nikita Bhattacharjee¹, Prathamesh Kalyankar², Ishwar Wankhede³, Shubham Agarwal⁴, Shridhar Jadhav⁵

Department of Computer Engineering
1,2,3,4,5 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

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

Mobile Marketing basically consists of three major Opportunities and Challenges. As indicated in Chandrasekar Vuppalapati,et.al [3], those Challenges are 1) The potential impact of Mobility in Digital Marketing, 2) The unprecedented adoption and 3) The customer engagement challenges due to huge mobile datasets. This paper has various algorithms such as Decision tree algorithm, K means Clustering algorithm and K nearest neighbor algorithm. Decision tree algorithm is popular technique for classification as no domain knowledge is needed for constructing the decision tree. This paper is overcoming the limitation of traditional Hadoop deployment by introducing an in-memory cluster computing platform which increases the performance by 100x times.

According to traffic forecast report of Cisco Systems half a billion mobiles were sold in 2015, and mobile traffic grew by 74 percent Shaowei Lin, et.al.,[4]. This paper shows the pros of deep learning which are, High Accuracy in Mobile Data Analytics and Utilization of huge unlabeled mobile data for unsupervised feature extraction. The limitation is the curse of dimensionality and size of Mobile Data. Learning deep models in this size of data takes time between few hours to several days in conventional computing systems. The pros of deep learning also include generation of intrinsic features

required in mobile data analysis. This paper shows two main steps for learning deep models 1) Gradient Computing and 2) Parameter Update. This paper shows that main role of Apache Spark Platform is to tackle the three v's volume, volatility and velocity aspect in data analysis.

Recently the huge data of mobile cellular networks hasn't been paid much attention Ying He [5]. This paper shows the extraction of insightful information through big data analysis. It also shows the capability of collection and scattering the data to understand user behavior, choices and priorities. This paper also shows that users time table and daily habits can be derived from the pattern of their usage of data. This paper shows representation of big data through random matrix theory. There is systematic exploration big data analytics and mobile cellular network. This paper shows some research challenges and big data analytics prospects for next generation cellular network.

In recent years, the enormous growth of mobile marketing has left users with large storage of data, Prakash V Parande, et.al., [6]. The problem statement of this paper is to find nearby hotels, colleges and hospitals by the given latitude and longitude range limit. The strict limitation of the existing system is that, while solving this problem the complexity increases as data volume increases. To overcome this limitation new matching concept called mapping is used.

This paper characterizes the computational social science in four dominant paradigms such as, 1) Text analysis, 2) Social network analysis, 3) Social complexity analysis, 4) Social simulations Ravi Vatrappu, et.al., [7]. This paper uses Sentiment analysis for calculating sentiments of artifacts and actors based on presented formal model. The limitation of this paper is that it doesn't discuss about visual analytics, crisis communication, crisis management, labour rights. It also includes limitation of limited space given to computational aspects of visual analytics tool. The future work of this paper is to extend social set analysis in terms of formal models and analytical methods to include rough and random sets.

The key element to build predictive model is the ability to capture historical data of user behaviour. Ying Li, et.al., [8]. This paper uses four different sampling methods such as, under-sampling, fixed oversampling, randomized oversampling and SMOTE-based sampling for managing unbalanced data. The proposed approach in this paper is exploration of data sampling mechanisms by trying different ratios of positive and negative result. Another exploration may be observing model performance changes with varying amount of data.

The promotion of mobile customers by data traffic is evaluated using decision tree and neural network data mining techniques, Huaying Shu, et.al. [9]. This Paper focuses on ANN Artificial Neural Network which is best suited in diagnostic and predictive analytic problems which mainly uses supervised learning technique. In future to promote the mobile data service the paper represents integration of different data mining methods to provide more detailed feedbacks to mobile operators. This paper shows the use of data mining techniques for the analysis of mobile customers on the basis of their specific characteristics and to synthetically analyse the consumers, Weng Suxiang, et.al., [10]. For the consumer characteristic analysis the data mining technique known as Clustering is used and for consumer buying mode analysis relating technique is used. It uses machine learning and statistics for establishing model of consumer behaviour. This paper shows the advantage of good capability and multiple classes of decision tree.

III. SYSTEM REQUIREMENTS

Hardware Requirements:-

- 3.2GHz Intel i3 processor
- 2 GB RAM

Software Requirements:-

- Apache Hadoop
- MYSQL Server
- XAMMP
- Eclipse IDE

IV. 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 their 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.

V. ACKNOWLEDGMENT

We wish to express our profound thanks to all who helped us directly or indirectly in making this paper. We also wish to thank all our friends and well-wishers who supported us in completing this paper successfully. We are especially grateful to our guide Prof. Nikita Bhattacharjee, for her time to

time, very much needed, valuable guidance. Without her full support and cheerful encouragement, the paper would not have been completed on time. We also wish to thank our director Dr. S.S Sonavane and HOD Prof. Soumitra Das for their kind support.

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 Vuppapapati, 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 Vatrappu, 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