

A Survey on Customer Behavior Analysis using Big Data Analytics

PRIYA G NAMBIAR¹, E K GIRISAN²

¹*MPhil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India*

²*Associate Professor, Department of Computer Science, Sree Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu, India*

Abstract-The lifestyle changes make the younger generation to use the online shopping approach drastically and this booms the business of E-commerce. The success of business is to know the requirement of the customer and provide sales at right time. Analyzing the behavior of customer is a crucial task to enhance the online shopping business. Big data is an emerging technology that can be used as a roadmap to predict the customer behavior and understand the preference by analyzing the data and transform it into valuable insights. Although there are many research and system studies have been carried out recently, big data has been proved to be more effective tool for customer behavior analytics. The main drawbacks of the previous studies are that they need accurate and more number of data set and features. This paper surveys about the recent data mining techniques applied for analyzing the customer behavior.

Keywords- Data visualization, Behavior Analysis, Data mining, Big Data Analytics, Big data.

I. INTRODUCTION

E-commerce has significantly changed the traditional buyer-seller relationship and as well as the shopping process for many customers. Customer behavior analysis is a most important and valuable task for improving online shopping business nowadays. On the other hand, study of user's behavior on e-commerce sites is not an easy task. Online shopping became complex as individuals spend large amount of data, therefore it has become more important to understand and characterize customer online purchase behavior. Recent days number of online shopping sites are rapidly growth due to more online shopping interest. These sites allow the user to buy an item based on their own willingness. Nowadays online shopping becomes extremely competitive because of the involvement of many vendors in the business. To improve online shopping, effective survey is needed, to find potential buyers based upon their transaction history and customer owned interest. The customer purchase behavior is analyzed based upon many important factors such as number of web page clicks, duration of the session and number of clicks per session and click through on more content etc. Customer can

take much different action, which significantly varying the level of effort.

Various data mining techniques such as Neural Networks, Decision Tress, Naïve Bayes, C4.5 classification algorithms are used to analyze the customer behavior. Decision tree algorithm is used to provide the classified report for the Behavior Analysis. The Naïve Bayes method is used to analysis and predicts the customer behavior through probability. The Neural Network provides the minimized error of the prediction of Customer Behavior. The C4.5 is often referred to as a statistical classifier. This is used to generate a tree based on customer buying behavior. In all this above mentioned techniques Customer buying Behavior are classified and predicted continuously. Most of the data mining techniques used nowadays have some limitation in point of view to data mining for an e-commerce application. Bigdata analytics can be used to leverage the power of data to fuel sales, so it makes a perfect sense to be used in marketing strategy by way of increasing the revenue and improving the brand image through customer behavioral analytics.

A. Big Data:

Big data is a collection of data that has very large volume, comes from variety of sources like web business organizations etc. in different formats and comes at us with a great velocity. Biggest challenge of bigdata is how to discover all the hidden information from huge amount of data collected from varied sources. There comes introduce customer behavior analysis

B. Customer Behavior Analysis:

Customer behavior analytics is the process of tracking, collecting and assessing of Customer data and activities from E-commerce site. This analysis mostly helps in marketing purpose which analyze and predict customer buying patterns. The Customer behavior analytics works based on collection of various data such as access information, web page clicks, transaction history past and current activity. Most of these activities helps in analyzing and predicting users' future activity. The important actions are list out below.

1. **Search:** Nearly every E-Commerce application effectively process with search functionality that allows user to discover previously unseen content. User may enter the searched data in E-commerce site and subsequently related data will be displayed in respond to search query.
2. **Close-up:** This application allows users to engage on-site with specific content to reveal more information and hence this process is called as close-up.
3. **Click-through:** This action allow user to engage off-side with content, that's why these actions are referred to as click through. Users can click on links to get the information from e-commerce site. This click through information helps to analyze the customer behavior more effectively.
4. **Save:** Many E- commerce sites allow user to book marks or save the content for later retrieval and viewing of the information
5. **Transaction history:** This is the most important and effective strategy. This action records all the information regarding searches and purchase made during each and every session. This action also maintains and categories the accessed information, so that it would be easier for Customer Behavior Analysis process.

These Customer analytics helps to turn big data into big value by allowing the E-Commerce site to predict the buyer behavior in an effective and fastest manner and thereby improving the sales, market optimization, inventory planning and can do much more.

II. RELATED WORKS

A. GIS Approach

An author in this paper [1] recommends that GIS is capable to integrate, to manage, and to analyze information regarding spatial information. GIS is separated by layer, are connected via their common representation framework. GIS differs from another information system with its own data indexing system. Data in GIS is managed according to their locations in space. GIS mapping approach uses color-coding to indicate customer behavior as it changes across geographic regions. A map is separated and divided into polygons that represent geographic regions that shows where churners are concentrated or where specific products are sold

B. Decision Tree based Models

The authors Nan-chen, Hsieh [2] describes Enhancing consumer behavior analysis by data mining techniques. This paper presents a two-stage framework of consumer behavior analytics first customer segmentation and customer profiling. This concept involves self-organizing map (SOM) neural network to divide customers into homogeneous groups of customers and a decision-tree simplified method to identifying

relevant knowledge. This concept gives the experimental results, and it shows that the proposed method is helpful for understanding the characteristic of customers and facilitates the development of marketing strategy.

Authors in [8] presented ID3 Decision Tree Algorithm for the forecast of Sales Prediction and Analysis from the big data series. Initially, mapped to different categories according to different values of the condition attribute set. All attributes are classified, i.e. discrete values. For each known value of the test attributes, a branch is created, based on which examples are divided. Following the same process, the decision tree of each division is generated recursively. Once an attribute appears in a node, it should not be considered again in any descendants of the node. The experiment helps the decision makers easily and effectively to sort out valuable associated information in the market analysis and also improves the sales performance in an efficient way.

From the paper [10] authors had presented decision-making model through the C5.0 decision tree. Classification is done based on the results of numerical calculation of some variables. The final result of the calculation is classified into several discrete values. The most important thing is boosting which give a weight to the sample, the higher the weight of the sample for the study generated the greater the impact of the decision tree. All the samples have the same weight in the initial state, each training sample weights should be adjusted in the next course, so that the error classification of a sample by the decision tree will obtain higher weight value at the next iteration. The experiments show that the accuracy of C5.0 algorithm was 86.65%, is significantly higher than other decision tree algorithm and it improves transaction success rate.

Author in the paper [13] has presented the customer behavior model to predict customer behavior by using the decision tree approaches. The common behaviors are centered by arranging the particular groups of customers in order to predict some behavior that how similar customer will behave under similar circumstances, by using the concept of customer behavior model, and it seek the behavioral information. This model predicts customer's behavior depends upon preparation of the data, segmentation and modeling.

Author in the paper [14] proposed that novel Tuned Decision Tree Predictive model to analyze the customer behavior in shopping. The recursive splitting method is used to build the tree. It has root node, internal nodes, and leaf nodes. The root node consists of input data. An internal node has many branches and is related to a decision. A leaf node presents the output of a supplied input vector. This concept introduces a model which is tuned by the Decision Tree Predictive model, that gives accurate results and therefore quality of prediction is certainly improved. The performance of this model is compared and validated with several online shopping data sets

also. It predicts the customer behavior and therefore it is used to optimize the business outcome.

C. Models based on Association Rule

Authors in the paper [3] proposed FP-Growth Association rule mining is introduced to detect relationship or affinity patterns across data and generates a set of rules. This algorithm works based on divide and conquer strategy. At first a transaction is read from the database. The algorithm checks whether the prefix of transaction maps to a path in the FP-Tree, to extract frequent item sets directly from this structure by using tree. Each and every node in a tree contains the label of an item along with a counter. After that, this shows the number of transactions mapped onto the given path. It structured a highly compact FP-tree, which is substantially smaller than the original database, by which costly database scans are saved in the subsequent mining processes. The pattern growth method is used to avoid the costly candidate generation. FP-growth is not able to find high utility item sets.

Authors in the paper [4] enriched the customer database with a prediction of customer's behavioral loyalty so as to develop it for target marketing. Cluster analysis and association rule mining model is used to identify the interesting customers based upon their purchasing history. After that segment-based knowledge inference method is used to find and derive the patterns like individual transition paths and to know the net path on which the customer is likely to move.

This paper [7] authors has presented the Association rules mining to detect hidden facts in the large market datasets and drawing inferences on how a subset of items influence the presence of another subset. The Association rules mining finds the interesting relations and connections along with large set of data items. The intention of this paper is to find frequent item set efficiently using Apriori Algorithm. This frequent sets are applied into the Boolean association rules. This algorithm is designed to operate on large database containing transaction. Initially a minimum support is applied to find all frequent item sets in a database after that frequent item sets and the minimum confidence constraint are used to form or generate rules.

Authors in the paper [11] proposed Closed + High Utility Item sets which combines with three proficient algorithms named such as THUP for finding high profitable item sets, TCFP for find out the closed frequent item sets and then, final output is get high utility of information. After, get the output, the item sets it will be maintained in a tree-based data structure named closed+ utility pattern tree (TCUP-Tree) such that the candidate item sets can be generated correctly and efficiently with only two scans of the database, then that will be segmented into the multiple clusters for speed computation. The CHUI directly computes the utility of item sets without producing the candidates.

D. Rough Set Approach

In the paper [6] authors proposed Rough Set Approach for analyzing customer needs and behaviors. This approach characterizes the behavior of customer to maintain long-term profitable customers. How to organize the behavior of customer is analyzed by choosing the important attributes such as customer intention from the database. The customers are then classified into groups according to their attribute values. This concept is used to predict the entrepreneur behavior The rough set approach focuses on customer segmentation and rule generation for customer behavior analysis. Experiments shows that clustering and induction algorithm is more efficient for analyzing the customer behavior.

E. Structure Time Series Model

The authors in paper [5] had presented the predictive study and Structure Time Series Model. The main intention of this paper is to effectively predict the clothing sales volume by combines web search data and structure time series model. Initially it elaborates the relationship between the search data and the sales volume, and builds a theoretical framework. Finally, this approach make use of e-commerce site search data combined with structure time series models to forecast the clothing sales volume, and also achieved a good prediction about it.

F. Hierarchical and Network Model

Authors in the paper [12] had presented and proposed Traditional Analytical Systems for Customer Behavior. This paper effectively analyzing customer behavior using two different models, hierarchical data model and network data model. At first these models are attempted to create a database standard and, then resolve some of the difficulties, such as its inability to represent complex relationships in the database management system. However, these two models had some fundamental disadvantages like complex program and minimal data independence.

G. Utility Pattern Growth Algorithm based Model

The authors in paper [9] had presented UP Growth which is one of the efficient algorithms. This newly introduced algorithm with a compact data structure is used for efficiently discovering the high utility item sets from transactional database. The information of transaction and item set, facilitates the mining performance and avoid scanning of original database redundantly which is maintained by the UP-Tree. UP-Tree (Utility Pattern Tree) is adopted, which scans database only twice to obtain candidate items and manage them in an efficient data structured way. This algorithm present in this paper is to generate high utility item sets efficiently for large data set and reduce the execution time.

H. Model based on Click Stream

Authors in the paper [15] proposed click stream data approach for Analysis and Prediction of customer behavior. Initially these click stream approaches extract information and make predictions about customers shopping behavior and also may provide some hints about their buying behavior. This model predicts whether customers will or will not buy their items are added to shopping baskets on a digital market place. The multi-layer neural network prediction data mining model is used for analysis. The click stream data approach effectively predicts the customer behavior with 90% accuracy.

Paper No.	Technique	Advantages	Disadvantages
1	GIS Model	This model is easy to implement with a large number of data set. Model accurately predicts where specific products sells the most	Time spent in a state is same for all states, so it takes more time.
2	Self-organizing map (SOM) And decision-tree	High performance which offers better accuracies for consumer behavior analysis	Computationally expensive
3	FP-Growth Algorithm	It is faster than other ARM Algorithm. Repeated database scan is eliminated	The memory consumption is more. Cant able to use for finding high utility item sets.
4	Cluster analysis and association rule Model	Efficient way of identifying the purchase interests of the customers	This method does not consider the data imbalance problem in the cluster task.

5	Structure Time Series Model (STSM)	Highly scalable, for clothing sales volume predictions. Achieve good prediction result.	This STSM model is expensive, both in terms of memory and compute time.
6	Rough Set Approach	Successful Model for analyzing customer needs and behavior.	Needs more feature. Feature extraction is more complicated
7	Apriori Algorithm	This algorithm has least memory Consumption. This will find frequent item set efficiently	It requires many scans of database. It allows only a single minimum support Threshold.8
8	ID3 Decision Tree	Effectively improve and forecast sales performance and prediction.	Only one attribute at a time is tested for making a decision. Does not handle missing values.
9	UP Growth Algorithm	Efficiently discovering high utility item sets from transactional database. This can apply for large data set.	The data base should be scanned every time for identifying the high utility and closed item sets.
10	C5.0 decision tree Algorithm	C5.0 can apply for big data set. It does improve transaction success rate.	Only 80% of accuracy can be achieved, lack of certain data security.

11	Closed + High Utility Itemsets Model	Algorithms reduce the number of candidates and database scans effectively. Reduce the communication cost.	The performance of the mining task decreases greatly for low minimum utility thresholds.
12	Hierarchical data and network data model	Extract and analyzing customer behavior is Accurate.	Complex program and minimal data independence
13	Decision tree	Effective customer behavior prediction. This is efficient even for massive data sets.	In some complex cases, splitting data into classes might not be helpful.
14	Tuned Decision Tree Predictive model	This model gives accurate result to analyze the customer behavior in shopping	it won't generate the reliable patterns for the case of noisy and missing data.
15.	click stream model	Predict the customer behavior with 90% accuracy.	Needs more parameters and deep study is necessary, so that can't able to apply big datasets

Table 1.0 Comparison table

The above table 1.0. Depicts the working methodologies of various data mining techniques which can be used to effectively achieve Customer Behavior Analysis.

III. CONCLUSION

Effective Customer Behavior Analysis is one of the major problems in nowadays which leads to many problems in shopping sites. Predicting and analysis behavior is possible only by the consideration of some of the important factors. This analyzing method of the factors can be achieved by the inclusion of data mining techniques. Data mining methodologies embraces methods such as Neural networks, Apriori, UP Growth, clustering mechanisms, classification, big data, etc. Further implementation has to be done in order to predict and analyze Customer Behavior in a big data environment.

IV. ACKNOWLEDGMENT

I wish to express my deep sense of gratitude to **almighty, my teachers, parents and my husband** who lead me throughout my studies. I wish to express my gratitude to my guide **E K GIRISAN**, Sree Krishna Adithya College of Arts and Science, Coimbatore for his excellent guidance, encouragement and great involvement in my work.

V. REFERENCES

- [1]. Murad, Abdulkader A. "Using customer's data for defining retail market size: a GIS approach." *International Journal of Business Information Systems* 4.3 (2009): 340-359.
- [2]. Hsieh, Nan-Chen, and Kuo-Chung Chu. "Enhancing consumer behavior analysis by data mining techniques." *International journal of information and management sciences* 20.1 (2009): 39-53.
- [3]. Bo Wu, Defu Zhang. "An Efficient Frequent Patterns Mining Algorithm Based on Apriori Algorithm and the FP-Tree Structure." *International Research Gate of Computer Applications* 108.16 (2010).
- [4]. Alias Devi, P. Isakki, and S. P. Rajagopalan. "Analysis of customer behavior using clustering and association rules." *International Journal of Computer Applications* (0975-8887) Volume (2012).
- [5]. Adhikari, Ratnadip, and R. K. Agrawal. "An introductory study on time series modeling and forecasting." *arXiv preprint arXiv:1302.6613* (2013).
- [6]. Dhandayudam, Prabha, and Ilango Krishnamurthi. "Customer behavior analysis using rough set approach." *Journal of theoretical and applied electronic commerce research* 8.2 (2013): 21-33.
- [7]. Hussain, RZ Inamul, and S. K. Srivatsa. "A Study of Different Association Rule Mining Techniques." *International Journal of Computer Applications* 108.16 (2014).
- [8]. Xurigan Saiyin, Chenna Hu, and Dou Tan1. "Research on Apparel Sales Forecast Based on ID3 Decision Tree Algorithm." *arXiv ICMII 2015* (2015).
- [9]. Xing, Shuning, et al. "Utility Pattern Mining Algorithm Based on Improved Utility Pattern Tree." *Computational Intelligence and Design (ISCID), 2015 8th International Symposium on*. Vol. 2. IEEE, 2015.
- [10]. Pandya, Rutvija, and Jayati Pandya. "C5. 0 algorithm to improved decision tree with feature selection and reduced error

- pruning." International Journal of Computer Applications 117.16 (2015): 18-21.
- [11].Joyce Jackson, Alva Erwin. " Combinatorial Temporal Closed+ High Utility Item set Mining Algorithm in Transactional Database" International Journal of Advanced Research in Computer and Communication Engineering.
- [12].Yuan, Baoli. "Applications and Analysis of Client Consumer Behavior Based on Big Data." Advanced Science and Technology Letters (2016).
- [13].Fokin, Dennis, and Joel Hagrot. "Constructing decision trees for user behavior prediction in the online consumer market." (2016).
- [14].Kalaivani Devi, T.Arunkumar. "Customer Behavior Predictive Modeling for Online Shopping using Tuned Decision Tree Method." International Journal of Computer Applications (01375–8887) Volume (2017).
- [15]. Pawar, Namrata. "Analysis and prediction of E-customers' behavior by mining clickstream data using Naive Bayes." (2018).