THE IMPACT OF CLICK STREAM ANALYSIS IN BUSINESS PROCESS

Dr. K.CHITRA

Assistant Professor, New Horizon College, Bangalore, India

Abstract - A clickstream is a list of all the pages viewed by a visitor, presented in the order the pages were viewed, also defined as the succession of mouse clicks that each visitor makes. A clickstream will show us when and where a person came in to a site, all the pages viewed, the time spent on each page, and when and where they left. An interactive click stream is a graphic representation of a click stream; a list of pages seen in the order in which they were visited. The graphic representation allows clicking on the pages, and seeing what the visitor saw, hence the label interactive. Web traffic analysis concentrates on how visitors move through the site. It measures the number of pages delivered to visitors, and determines how often visitors hit their browser stop button, how much of the page was delivered when they hit it and how long they waited before stopping the page.   For e-business analysis the clickstream data is used to determine higher-level information, tracking visitors' responses to pages and the content rather than on how they navigated that site. The main objective of this paper is to find out the impact of click stream analysis in business.

Keywords – Data mining, Web mining, Click stream analysis, Knowledge Discovery, Clustering, Data warehousing, Data Stream Analysis.

I. INTRODUCTION

Knowledge discovery in databases (KDD) is the field that is evolving to provide automated analysis solutions.   The most obvious reason for examining clickstreams is to extract specific information about what people are doing on the website. Examining individual clickstreams will give us the information we need to make content-related decisions without guessing. There is a wealth of information to be analyzed, we can examine visitor clickstreams in conjunction with any of the information provided by a good stats program: visit durations, search terms, ISPs, countries, browsers, etc. The process will give us insight into what the visitors are thinking.

The growth in the use of the Internet for Business to Customer (B2C), Customer to Business (C2B) and Business to Business (B2B) means the amount of data to be stored, managed is to be increased dramatically. Web data click volumes will increase by a factor of ten for every twelve months. Companies will need to build massive stores of this rich granular click-stream data.' This data will achieve terabyte proportions very quickly and will enable predictive modeling for content management, campaigns and personalization (potentially one to one marketing). All predicated on good data being collected, cleansed, stored and sensibly applied to real business opportunities.

Clickstream analysis will give us the opportunity to identify the search terms that are the most valuable for our website, by actually telling us how they perform. Clickstream analysis will tell us where they come in, what they look at, and where they leave. It is up to you to figure out why they leave (also known as shopping cart abandonment).

The section II shows the related work based on the concept of Data mining, Web mining, Data warehousing, Data stream mining, Clustering, Click stream analysis. The section III deals with the impact of click stream analysis in business. Section IV deals with the sample business analysis using click stream analysis. Section V describes the conclusion of this paper and References deals with the various reference papers available in various journals.

II. RELATED WORK

This Section deals the related work based on the concept of Data mining, web mining, Data warehousing, Data stream mining, Clustering, Click stream analysis. Phimphaka Taninpong et al.,[1] proposed a new text clustering algorithm based on tree structure. The proposed clustering algorithm is used to search all the clusters by traversing down the tree. The objective of this algorithm was to provide document clusters having high cohesion, and to keep the minimum number of clusters. The proposed clustering algorithm was used to spam mail classification. Amit Kumar Saxena et al., [2] proposed two hybrid methods which are the combination of filter-based feature selection, genetic algorithm, and sequential random search methods. The first proposed method is hybridization of information gain and genetic algorithm. Genetic algorithms are used to select optimal feature subset. The second model is the hybridization of information gain and sequential random K-nearest neighbor. Sunita Soni et al., [3] proposed a new Fuzzy Weighted Associative Classifier (FWAC) that generates classification rules using Fuzzy Weighted Support and Confidence framework. They proposed a theoretical model to introduce new associative classifier that takes advantage of Fuzzy Weighted Association rule mining. S Pramod et al., [4] discussed different type windowing techniques and the important algorithms available in this mining process. These algorithms are used for mining the data streams. They developed association rule mining algorithms for data streams. Bharat Singh, et al., [5] explained a new algorithm to deal the problem of outlier detection in High Dimensional data

with the help of descriptive analysis. The technique used in this algorithm is hybridization of density-based outlier detection and distance-based outlier detection technique. S Pramod, OP Vyas [6] developed a new algorithm to improve the performance of the online stream mining algorithm. The result shows that the new approach improved the performance in the association rule mining in its online environment. The implementation of this algorithm has been tested using the datasets. Kavita Das et al., [7] discussed about the primary techniques of web mining -Association rule mining, Classification rule mining and Clustering. There is necessary to understand the structure and behavior of surfers on the web, so that relevant pages can be quickly rendered to a surfer and make the browsing task more users. S Shrivastava et al., [8] provided an overview of data preparation techniques and algorithms that can be used in order to convert raw log data for click stream data mining. The data in the log files about the actions of the users can not generally be used for mining as it is stored. Sangita Babu & C.Jothi Venkateswaran [9] observed six working loads in the ten types of learning system. This paper attempted to observe the cognitive load such as mental, physical, temporal, performance, effort and frustration in the long term, short term, working, instant, responsive, process, recollect, reference, instruction and action memory and classify the observed values as per the generalized and specialized properties.

III. THE IMPACT OF CLICK STREAM ANALYSIS IN BUSINESS

This section describes about the uses of click stream data in the growth of business. The Click stream analysis also helps the clients to enhance their business by tracking various events generated by the users and also useful for the marketing purposes. The click stream data is also used for analytical purposes require analysts to prove or disprove a hypotheses (the deductive approach) that produce reports that can be acted upon.

Specific techniques that can be used for clustering include standard statistics, memory-based reasoning, neural networks, and decision trees. Specific techniques that can be used for estimating and predicting include standard statistics and neural networks for numerical variables, as well as all the techniques described for classifying when only predicting a discrete outcome.

The latest source of large amounts of data is the Internet clickstream. Smart sites record user's activity within 'the clickstream' this data includes where visitors went, how they got there, what they did and how long they stayed. As with all data it needs to be captured, cleansed, stored, analyzed and used as the basis of actionable decisions.

For e-business analysis the clickstream data is used to determine higher-level information, tracking visitors' responses to pages and the content rather than on how they navigated that site. Clickstream data gives the opportunity to identify the search terms that are the most valuable for decision making, by actually telling how the changes may happen and that will change the business process.

3.1 THE BUSINESS OBJECTIVE

The following are the two primary objectives – (i) Maintain a revenue/profit stream by meeting the needs of the sales and Marketing functions of the lenders through services - BPO. (ii) Develop an initial business relationship with lenders that would help ABC Corporation establish its credibility and showcase its other offerings. Establishing this relationship is easier with program because this is low risk and low integration service with a shorter sales cycle than the core BPO Service. In this paper, two business processes – Home price website, Email marketing are considered as example.

IV. THE SAMPLE BUSINESS PROCESS USING CLICK STREAM ANALYSIS

4.1 HOME PRICE WEBSITE:

The website associated with the Home price tool (sometimes called "Deed Register") to the borrowers/customers and provides links to the borrowers to the client Mortgage Banks to offer lending support/service. The website is offered free to consumers/borrowers and allows them to search for comparable homes sale prices in the specified geographical area. The website provides limited information related to the home the user is searching like living rooms, rest rooms, etc.   Registration is a critical source of predictive data on a user and it is provided by the user while he or she login to the home price website.
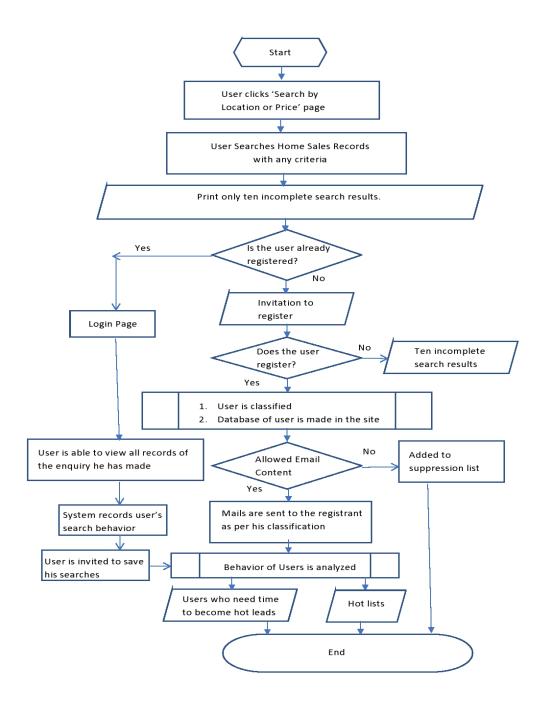
Fig 1. Process flow of Home price website

4.2 Email marketing:

The Email marketing program is made up of a set of email messages referred as "Grams". All registrant data is posted to the appropriate account in "Response account" for use in the Email Marketing Program.   All email messages, before reaching borrower, must be reviewed and approved by the client-lender for the contents and the look and feel nature of web based email messaging - 'before' sent to the customers / borrowers. ABC Corporation composes the proposed attractive emails to the borrowers in a separate environment and put them for the review of the clients (banks/ lenders) for the mails information approval.

The email marketing messages rely heavily on information collected at registration. Email messages are tailored specifically to three consumer variables that are given below to make decision by doing further analytics on them. The consumer variables are the loan product, the level of urgency of the consumer, the segmentation of the user (first time buyer, move up buyer)

V. CONCLUSION

The latest source of large amounts of data in the Internet is clickstream. Smart sites record user's activity within 'the clickstream'. This data includes where visitors went, how they got there, what they did and how long they stayed. As with all data it needs to be captured, cleansed, stored, analyzed and used as the basis of actionable decisions. Clickstream data gives the opportunity to identify the search terms that are the most valuable for decision making, by actually telling how the changes may happen and that will change the business process.

REFERENCES

[1] Phimphaka Taninpong,Sudsanguan Ngamsuriyaroj ,"Tree-based text stream clustering with application to spam mail classification", International Journal of Data Mining, Modelling and Management (IJDMMM), Vol. 10, No. 4, 2018.

[2] Amit Kumar Saxena, Vimal Kumar Dubey, John Wang, "Hybrid feature selection methods for high-dimensional multi-class datasets", International Journal of Data Mining, Modelling and Management (IJDMMM), Vol. 9, No. 4, 2017.

[3] Sunita Soni, OP Vyas, "Fuzzy Weighted Associative Classifier: A Predictive Technique For Health Care Data Mining", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Volume 2, Issue 1, Pages 11-22.

[4] S Pramod, OP Vyas, " Data stream mining: A review on windowing approach", Global Journal of Computer Science and Technology Software & Data Engineering, Volume 12,Issue 11,Pages 26-30.

[5] Bharat Singh, Nidhi Kushwaha, OP Vyas, "Exploiting the anomaly detection for high dimensional data using descriptive approach of data mining", 4th International Conference on Computer and Communication Technology (ICCCT) 2013,Pages 121-128.

[6] S Pramod, OP Vyas," Recent frequent item sets mining over data streams" ,Conference Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology, Pages 484-489, Publisher ACM.

[7] Kavita Das, OP Vyas, Clemens H Cap, Anne Gutschmidt, "Suitability of Web Usage Mining for Web Content Syndication", Proceedings of National Conference, Pages 26-27.

[8] S Shrivastava, OP Vyas, CH Cap, A Gutschmidt, "Preprocessing", Proceedings of the 3 rd National Conference - INDIACom-2009 Computing For Nation Development, February 26 – 27, 2009, Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi.

[9] Sangita Babu, C.Jothi Venkateswaran, "Cognitive Load Persuade Attribute for Special Need Education System Using Data Mining", International Journal of Data Mining Techniques and Applications (IJDMTA) ISSN: 2278-2419 pages 4-8.