International Journal of Business
Management & Research (IJBMR)
ISSN(P): 2249-6920; ISSN(E): 2249-8036
Special Edition, Jan 2015, 371-376

CLICKSTREAM DATA VISUALIZATION FOR E-COMMERCE WEBSITES

SARA S. SAMSON

University of Michigan, Dearborn, Michigan, United States

ABSTRACT

Internet is the trend word of past 3 decades and the power of internet has open wide angles for various markets. One such booming market is E-Commerce. The biggest issue faced by data scientists is to create value and make decisions based on the enormous amount of data that flows in every day. Click stream data is the data that is generated as and when a user clicks on the web page. This data helps to demonstrate the behavioral pattern of a user and much more vital information. In this paper, the data visualization of click stream data for an e-commerce website is demonstrated using SAP lumira. Various techniques to perform data visualization are also discussed.

KEYWORDS: Click Stream Data, Data Visualization, E-Commerce

INTRODUCTION

The evolution of high speed internet and the increase in the number of internet users every day has paved the way to the boom of e-commerce platforms. E-commerce is preferred over brick and mortar retails due to its vast availability of products, convenience and fast transactions [9]. With the increase in the number of e-commerce portals, comes the increase in the competition among the e-commerce firms and in-flow of data by the users. Competitive edge could be achieved only when the firm could make meaningful insights from the humungous amount of data that flows in every day. However, obtaining knowledge from this data is not an easy task. The data includes structured and unstructured data from different environments. Marketing managers strive hard to form behavioral patterns of users visiting their websites every day to cater to their change in needs, choices and change in preferences by using various data mining technologies [6]. Data mining is the process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequent knowledge from large databases [1]. When data mining is done on web streaming through documents, http logs, has tags, app server logs, it is called as web mining [9].

- **Web Mining**

When a customer visits a site to look for information, browse through catalogs of products or checks out and makes a transaction, the customer leaves behind trails of data as the source for web data mining. This data could be used to identify the customer's behavioral patterns, their preference, needs and wants. Another source of data is the product information present in the website such as price, uses, durability etc., which are presented to the customers. The structure in which they are placed could also serve as a data for analysis. Web data mining could be broadly classified into three categories:

- Web Content Mining
- Web Structure Mining

- Web Usage Mining

- **Web Content Mining**

As the name implies, web content mining is the process of obtaining potential, viable, valuable knowledge or model from the web contents present in the catalog description, information which details about a product, search results on the web page etc., In short, web content mining is a combination of text and multi-media mining [5]. Web content mining is used to determine knowledge from the content present in the web from the exact location it is placed and bring out hidden link or relationship between the data.

- **Web Structure Mining**

Web structure mining includes hyperlink mining, URL mining and internal mining. In an e-commerce website, the useful information can not only be retrieved from the content present in it but also from the way the website is structured and how each hyperlink navigates through each other. If a hyperlink is pointed more than often, then the structure of the webpage could be re-structured to make the hyperlink more visible and easy to navigate [4].

- **Web Usage Mining**

Web usage mining could be personalized service pattern mining or general access pattern mining. Typically when a customer navigates into an e-commerce website, they generate a pattern in which they either navigate to the webpage like through a promotional advertisement in a search engine or third party website or through directly typing in the URL bar. When the customer enters the webpage, they either browse through it generally or make a specific purchase. This is recorded as the customer behavioral pattern by accessing the app server logs, web logs, the search logs etc.

DATA PRE-PROCESSING

Once the data is collected pre-processing, pattern discovery and pattern analysis is carried out. Various algorithms are carried out to execute the processing of data. By doing this, we could eliminate the unwanted data, redundant data, missing data etc. The data is later grouped into user session with the help of URL suffix; cookie etc. so that clustering of data is possible [8].

- **Pattern Discovery**

Once the data is pre-possessed and grouped, pattern discovery could be done. The most common techniques of pattern discovery for web transactions are:

- **Path Analysis:** It is the simplest technique with which the graphical representation of a website with the webpage as nodes and hypertext as links form a path in the webpage. The most travelled path, the entry and exit point, the duration of a session using the web log data is identified using this technique.
- **Association Rule:** Association rule is typically used for market based analysis to associate products that are brought together by which the user behavioral buying patterns could be formulated. This could help in formulating effective market strategies.
- **Sequential Patterns:** By obtaining sequential patterns, one could identify the sequence of items viewed over a particular time stamp and helps in targeting advertising aimed at users or user groups.

- **Clustering**: Clustering is done mainly to group an item set based on nodal information, such as the ease of data retrieval is achieved in a much faster pace.

CLICKSTREAM ANALYSIS

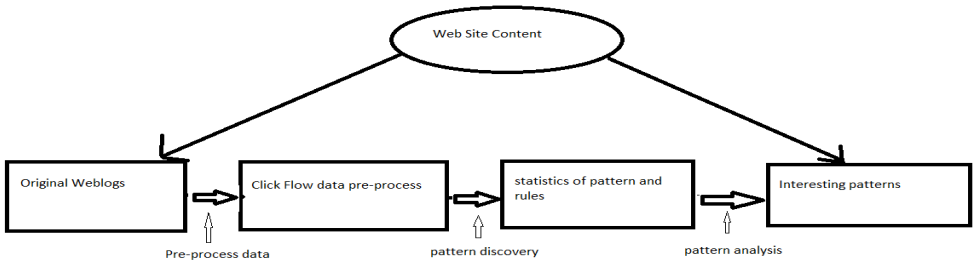
A clickstream is the recording of the parts of the screen a computer user clicks on while web browsing or using another software application. As the user clicks anywhere in the webpage or application, the action is logged on a client or inside the web server, as well as possibly the web browser, router, proxy server or ad server. By using clickstream analysis we could derive data from various parameters such as,

- **Visit Duration:** the duration of time a customer spend during one session
- **Time Stamp:** the time and date variables of an active session by a customer
- **IP Address:** the IP address from which the customer viewed the webpage
- **Session Details:** the number of links clicked and navigated, the items viewed, transactions made if any etc.,
- **Product Details:** The sequence of the products viewed and the number of repetition of the products viewed etc.

Clickstream analysis is useful for web activity analysis. A customer may visit a website just for browsing over products or specific product purchase. Knowing different user behavior patterns helps conduct market segmentations in order to develop marketing strategies, or enhance personalized shopping experience. In practice, however, learning the various user behavior patterns is nontrivial [10]. The most striking questions to enhance sales and customer retention in an e-commerce website would be,

- What are the most frequent user behavior patterns?
- What are the demographics of the users who follow a specific behavior pattern?
- How do the behavior patterns correlate with the performance of the online service?

To sum them up in a single image,



Source: Retrieved from application study from web-mining in e-commerce [5].

Figure 1

VISUALIZATION OF CLICKSTREAM DATA

Clickstream Data could be clustered and mapped in lower levels using algorithms like Self-organizing maps [10], accelerated ants algorithm [3] etc. But real time analytics of live data require handling volumes of data to give a visual output. Visualization of data is very important in present day scenario as it helps to improve understanding, maintain focus,

to tackle the growing volume of data and to improve the decision making ability. Click stream data could be visualized by assigning shapes and colors to each user session depending on the parameters present in it. The proportion of the shape or color could vary based on the frequency of the pattern. These symbols could be placed on a plotting map to arrive on a pattern. However, the overlapping of the symbols could be a major problem to be sought out. This overlapping of symbols could be eradicated by the use of algorithms such as k-means algorithm, SOM with Markov chain model etc.

Let us consider the real world example of data visualization with click stream data using IBM lumira application.

The workspace area of IBM lumira consist of dimensions and measures on the left panel with filters on the right and workspace in the middle as shown in figure 2a. To demonstrate the visualization of clickstream data, a dataset of e-commerce web site is loaded into the system. With the slice and dice options, we have narrowed down the sales revenue and gross margin or number of short sleeve shirts sold in Dallas

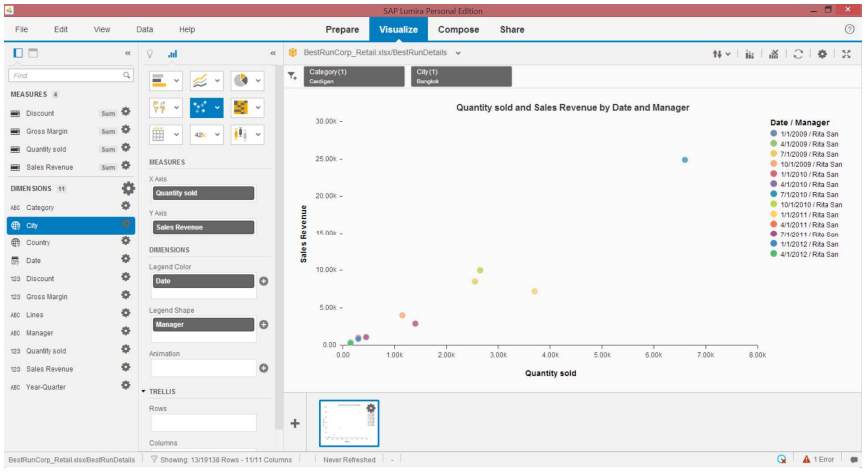


Figure 2(a): The Workspace Area of SAP Lumira Loaded with the Clickstream Data Set

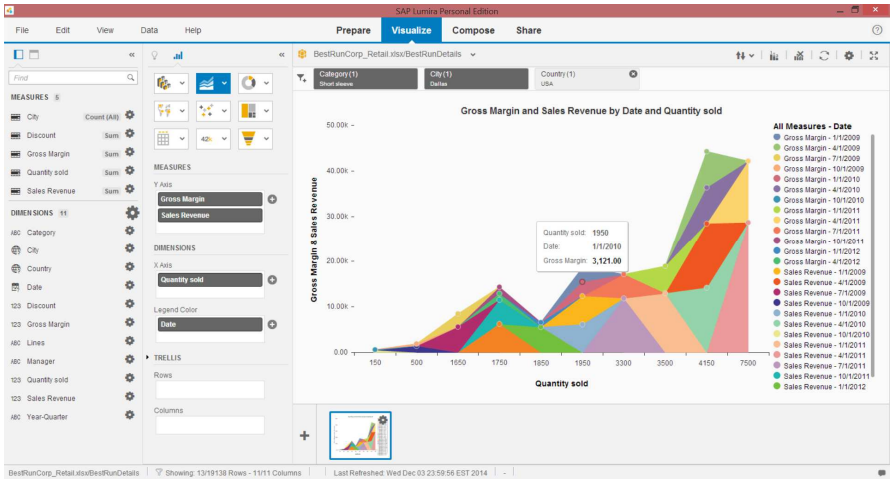


Figure 2(b): The Data is Filtered with the Quantity of Short-sleeve Shirts Sold in Dallas against the Sales Revenue and Gross Margin for Each Day Indicated by Various Legend Colors

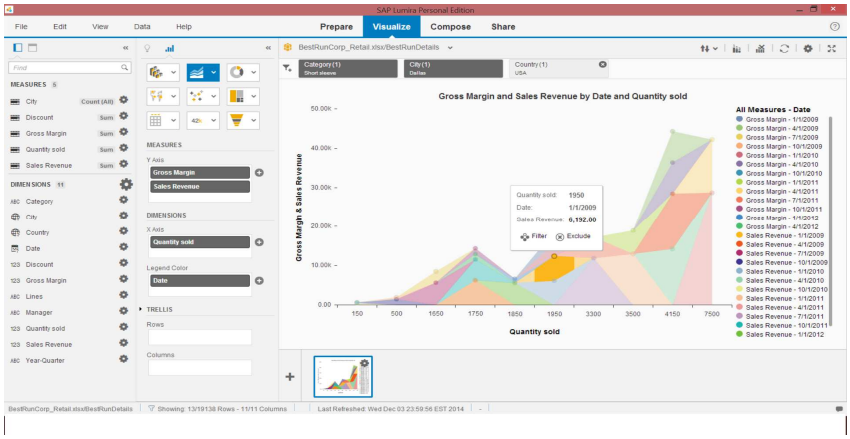


Figure 2(c): The Data is Further Sliced to a Particular Date by Selecting the Portion and Clicking on the Filter Option

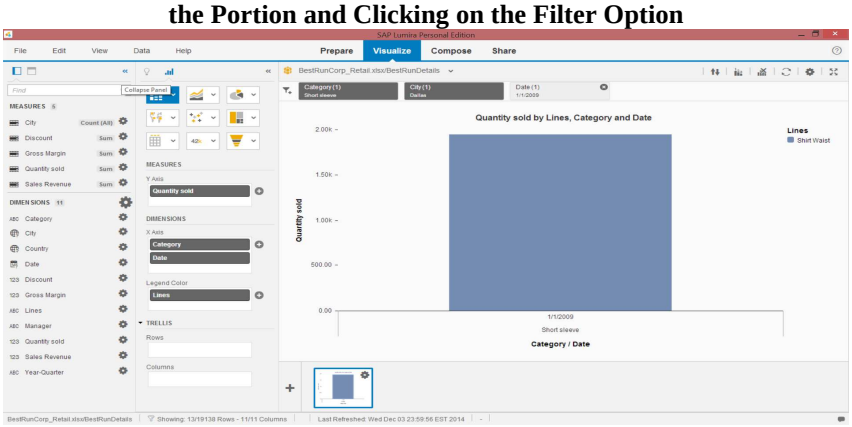


Figure 2(d): The Selected Portion is further Filtered to find the Quantity of Short Sleeve Shirts sold on a Particular Day in Dallas, USA

Future we could prepare, visualize and compose reports with the available click stream data.

DATA VISUALIZATION TECHNIQUES

Scatter plots, scatter plot matrices, grand tour and data tour, bar charts, funnel charts, pie-charts, heap map, geographic charts etc., are a few visualization methods to align or display a large set of data [11]. With these visualization tools, various techniques could be used to ramp down the data and narrow on particular criteria. Few of those techniques are:

Funnel Analysis: In funnel analysis, we determine the entry and exit point and take a set of user sessions data. When demonstrated in a funnel map, the analysis would show how many users have completed through the entire session and how many have dropped down on particular stages. E.g. when we set entry and checkout as the entry and exit criteria for a funnel analysis, we could determine how many users have made a purchase and how many have dropped out of it [2].

Path Analysis: This again sets two points as entry and exit points and takes data sets of user sessions. This analysis helps to determine different paths take by a user to reach the exit point from the entry path. This helps in finding the most favored user path [2].

Decision Trees: Decision trees are used to subset data till no further subset could be reached. This is one of the most efficient methods used for predictive analysis [7]. Decision trees could be visualized with criteria to eliminate unwanted information and narrow down on customer segments to target on marketing campaigns.

CONCLUSIONS

Pictures speak volumes. Data visualization is an effective tool for managers to make important decision. Data visualization helps both in predictive and descriptive analysis by providing insights on data which gives a meaningful perspective of a scenario. Data visualization is not a new technology. But the correct usage of the tool in any industry with the right data is crucial. In markets such as E-commerce which counts on volumes of data every day, analysis of it is necessary to stay and succeed in business. This paper gives a brief introduction of data visualization and is restricted in reviewing to only one tool, SAP lumira with clickstream data. It could further be extended to various other tools in market and analyze the performance of each tool in different industries with different data.

REFERENCES

1. Delen, Sharda, Turban, "*Business Intelligence: A Managerial Perspective on Analysis*", third edition (2014), New Jersey, Pearson.
2. Alan Keahey T, Eick G Stephen(2002), "Visual Path Analysis" *Proceedings of the IEEE Symposium on Information Visualization*(InfoVis'02), 165 – 168, DOI:10.1109/INFVIS.2002.1173163
3. Hannah Inbarani H, Thangavel K (2006), "Click stream Intelligent Clustering Using Accelerated Ant Colony Algorithm" *IEEE International Conference on Advanced Computing and Communications, (ADCOM)* . 129-134, DOI: 10.1109/ADCOM.2006.4289869.
4. Hsien Ting(2008) "Web Mining Techniques for On-line Social Networks Analysis" *IEEE International Conference on* Service Systems and Service Management, DOI:10.1109/ICSSSM.2008.4598506
5. Li Haigang Yin wanling(2006), "Study of Application of Web Mining Techniques in E-Business" *IEEE International Conference on* Service Systems and Service Management, 2,1587 - 1592 DOI: 10.1109/ICSSSM.2006.320773
6. Liu Yanbin, Yuan Ping (2010)"Processing Online Market Information based on Users Online Information Behavior" *IEEE Third International Symposium on Information Processing.58-62,*DOI:10.1109/ISIP.2010.84
7. Todd Barlow, Padraic Neville(2001) "Case Study: Visualization for Decision Tree Analysis in Data Mining" *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS'01), 149,* DOI:10.1109/INFVIS.2001.963292.
8. Yang Haijing(2006) "Study on Book Purchase Strategy Based on Web Data Mining Technology" *American Journal* of *Orthopsychiatry*. *76*(4), *482–488.* DOI*:* 10.1037/0002-9432.76.4.482.
9. Yonghua Zhao, Hong Lin(2014) "Web data mining applications in e-commerce" *The 9th International Conference on Computer Science & Education (ICCSE 2014),*557-559, DOI: 10.1109/ICCSE.2014.6926523
10. Zeqian Shen Jishang Wei et al(2012) "Visual Cluster Exploration of Web Click stream Data" *IEEE Symposium on Visual Analytics Science and Technology*, Seattle,65-72, DOI: 10.1109/LDAV.2012.6378977
11. Fang Fang Zhang (2008) "The Application of Visualization Technology on E-commerce Data Mining" *Second international symposium on intelligent technology application,* Shanghai, 563-566, DOI 10.1109/IITA.2008.18.