

Study & Analysis of Web Usage to Identify Customers' Behavior for Web Personalization

Prof. Narendrakumar S. Patel
U.V.Patel College of Engineering
Ganpat University, Kherva
Gujarat, India

Dr. Narendra J. Patel
U.V.Patel College of Engineering
Ganpat University, Kherva
Gujarat, India

Dr. Ashok R. Patel
Director, Dept. of Comp. Sci.
H. North Gujarat University, Patan
Gujarat, India

Abstract: In today's era, identifying interest of potential customer is utmost important for any business. A website usage data of any company provides very important details about the potential customers of the company. Interestingness of the customers can be analyzed well using various Web Mining techniques. The web usage analysis of total impressions and clicks on the selected pages of the site gives idea about popularity of products. The company can plan future business campaigns based on this data. The analysis also helps for better customization of the site for web personalization. Discrimination of various customer groups and their usage is analyzed. The data discrimination is performed at various levels and the promotion plans are designed as per the technical analysis.

Keywords: *Web Mining, Impressions, Clicks, CPC, CPM, Discrimination, Web Personalization*

I. INTRODUCTION

In this paper, data of a website usage is used to find out how many users click on the website from total number of impressions. This study is carried out on website of an engineering company Moris Industries. Website www.morisequipment.com is given for AdSense Google advertisement. Data is collected from the advertisement agency that is monitoring the transactions on the website. The data is analyzed by classifying it State wise and then applying various Web Mining analytical functions. We have identified useful dimensions like Region, Impressions, Clicks, CPC and CPM. It can be helpful to the owner of the company to find out customers scope in various states of India. With the help of comparative analysis of usage of different states it becomes easy to find out the core areas to concentrate on and the areas where people have more interest in knowing about the products of the company. The comparison of Impressions and Clicks is very important which gives idea about product popularity, restructuring and redesigning of the site.

II. DIMENSIONS ANALYZED

A. IMPRESSIONS

The exact number of frequency a specific Web site has been viewed or accessed by a web user. A page impression acts as a counter for Web pages informing website owners how many numbers of times their websites were visited [4].

B. CLICKS

Clicks refer to total number of hits on a web page or a specific web part on it. This count is required for further analysis of web page.

C. CPC

Cost per click (CPC) (also referred as a Pay per click) is an online marketing and advertising model used to direct traffic to websites, where advertisers pay to the publisher (that is a website owner) when the web part of advertisement is clicked by the web user [4]. The Google Adwords is one of the popular models that works based upon CPC concepts. Generally search and text advertising is sold by CPC model. In such type of advertising and marketing model website owners just pay for number of clicks and website owners get on their ads irrespective of number of impressions it takes to generate those clicks. For instance, if the CPC value is \$4.00 and your advertisement is publicized 12,000 times but gets no clicks then you pay nothing. If you get 10 clicks on your advertisement then you pay $\$4.00 \times 10 = \40.00 . No clicks, no revenue. It's a very simple formula.

$$\text{CPC} = \text{Total Cost} / \text{Total Clicks.}$$

$$\text{Total Cost} = \text{CPC} * \text{Total Clicks}$$

$$\text{CPC} = ((\text{Total Impression} * \text{CPM}) / (1000 * \text{Clicks}))$$

This is how much you would pay the ad-network or website every time a visitor clicks on your banner. CPC rates can be as high as \$3 per click or as little as 5 cents per click. It depends on your product and your market – amongst other factors, the more competition there is – the higher you will probably end up paying as you compete with competitors. This approach is differing from the "pay per impression" techniques applied in newspaper and television advertising.

D. CPM

CPM referred as Cost per 1000 Impressions (i.e. how many number of times the advertisement is publicized) (M is Roman numeral for 1000). Generally show advertising (e.g. images, banners, animations) is sold in CPM [4]. If the advertisement is exposed 1000 times then the cost will be equal to 1 CPM price. For instance, if a publisher charges \$20 CPM that means you will be shown 1000 times for \$20. If your budget is say

\$20,000 then mean your ad will be shown 1,000,000 times ($\$20,000 * (1000/\$20)$).

$$\text{Total Impressions} = (\text{Total Cost or Budget}) * (1000/\text{CPM})$$

$$\text{Total Cost} = (\text{Total Impressions} * \text{CPM})/1000$$

$$\text{CPM} = (\text{CPC} * \text{clicks} * 1000) / \text{Total Impressions}$$

Cost per impression often referred as a CPM (Cost per mile) or CPI is terms used in internet advertising and marketing related to web traffic. They refer to the cost of online marketing campaigns where advertisers pay for every time their advertisement is displayed, usually in the form of a banner advertisement on a website, but can also refer to advertisements in Email advertising. A single web page may contain multiple advertisements. In such scenarios, a single page view would result in one impression for each advertisement presented. In order to count the impressions served as accurately as possible and prevent fraud, an advertisement server may exclude certain non-qualifying activities such as page-refreshes or other user actions from counting as impressions. When advertising charges are expressed as CPM or CPI, this is the amount paid for every thousand qualifying impressions served. CPM advertising is often preferred by publishers because they can be more certain about the revenue they will generate from their website traffic. Cost per mille is the closest online advertising strategy to those offered in other media such as television or print, which sell advertising based on estimated viewership or readership. CPM provides a comparable measure to contrast internet advertising with other media.

III. WEB MINING TECHNIQUES

Expansion of the World Wide Web (Web for short) is a high volume of data that normally now freely available to the user's access has been. Managing and organizing different types of data that they can be accessed efficiently. Therefore, the use of Web data mining techniques now is the focus of a growing number of researchers.

Many data mining methods are discovering information hidden in the Web. However, Web mining does not only mean applying data mining methods and techniques to the data available in the Web. Algorithms to be customized to better suit the demands of the web. New methods and techniques better matched to the properties of web data should be used. In addition, data mining is not only algorithms, but also artificial intelligence, information retrieval and natural language processing techniques can be used efficiently. Thus, Web mining has evolved into an autonomous research. Web mining applications that store data on the web in search and extract hidden information consist of a wide range of reason. Another important mean of Web mining is to make data more efficiently and to offer adequate access mechanism. The third interesting approach is the information that users' activities, the log file of web access, for instance, are stored for predictive web caching can be obtained from the search. Thus, the web part of web mining is to be mined by three different sets can be classified. These are three classes (I) Web content mining, (II) Web structure mining and (III) Web usage mining. [3]

Web usage mining is the technique of discovering those activities while they are navigating through the web browsing.

Navigation references of visitors for the purpose of understanding the quality of electronic commerce services (i.e. ecommerce), to personalize the website or web portal or web structure is to enhance and improve the performance of web servers

Class Comparison

Class Comparison is used for Data discrimination. It generates what are called *discriminant rules* and is basically the comparison of the common attributes of objects between two classes referred to as the *target class* and the *contrasting class*. The methods used for data discrimination are very similar to the methods used for data characterization with the exception that data discrimination results include comparative measures.

IV. DATA PREPROCESSING

To analyze usage data is collected from usage log of www.morisequipment.com. This data is stored in database. Organize state wise with summarized values of each field. Additional columns like Standard GMT, Campaign, Website, etc removed from the data and only the columns to analyze State, Impression, Clicks, Avg CPC, Avg CPM is kept for further processing. It was secondary data, but data was not cleaned. So, the data is carefully cleaned and prepared for further analysis.

Here the date wise values of clicks and impressions are aggregated state wise. Based upon the interestingness of the attributes generalization is performed and the Prime Generalize Relation is obtained using attribute oriented induction. Now the data was processed and ready to use. The data on which the analysis is applied is given below.

TABLE I. PRIME GENERALIZED RELATION

Region	Impressions	Clicks	Average CPC	Avg CPM
All Other Regions	90.65	23.46	273.05	41.45
Andhra Pradesh	329.40	108.57	561.89	370.52
Assam	80.77	31.82	545.86	650.95
Bihar	87.88	40.63	671.56	845.98
Chhattisgarh	4.75	50.00	915.75	1119.43
Delhi	504.77	102.74	705.84	300.71
Gujarat	258.27	114.67	672.31	194.26
Haryana	159.29	51.43	466.57	439.00
Jammu and Kashmir	21.38	62.50	997.38	2806.58
Karnataka	558.09	168.83	679.82	248.01
Kerala	559.68	140.00	557.64	338.03
Madhya Pradesh	159.97	50.00	402.37	253.78
Maharashtra	734.88	242.11	1000.64	201.46
Orissa	99.34	37.93	462.28	529.88
Pondicherry	294.33	100.00	3059.67	102.73
Punjab	179.78	46.30	554.57	246.44
Rajasthan	135.00	55.88	464.15	801.63
Tamil Nadu	786.61	225.93	715.00	154.92
Uttar Pradesh	223.52	74.19	520.69	292.24
West Bengal	255.24	70.91	504.73	304.69

V. APPLYING DISCRIMINATION

Discrimination rule is applied on selected states with major difference. Table II gives details about the states where total number of impressions and clicks has vast difference among impressions and clicks. In some regions count of average clicks is very less in comparison to count of average impressions.

TABLE II. REGION WISE IMPRESSIONS AND CLICKS COMPARISON

Region	Chhattisgarh	Gujarat	Kerala	Tamil Nadu	Total Selected Regions
Impressions	4.75	258.27	559.68	786.61	1609.31
Clicks	50.00	114.67	140.00	225.93	530.59
Average CPC	915.75	672.31	557.64	715.00	2860.70
Avg CPM	1119.43	194.26	338.03	154.92	1806.64

The rule for quantitative discrimination is given in (1) and can be represented by associating the corresponding t-weight value with each disjunct covering the target class. The value of t-weight is [0.0, 1.0] or [0%, 100%]. Here two different kinds of values Impressions and Clicks of each state is compared for identifying difference between them.

$$\forall X, t \text{ arg et_class}(X) \Rightarrow \text{condition}_1(X)[t : w_1]$$

$$\vee \dots \vee \text{Condition}_m(X)[t : w_m]$$

(1)

Applying the quantitative discrimination rule to the data given in TABLE II, we get following rules. The t-weight value of Impressions and Clicks indicates total number of counts for each.

$$\forall X, \text{Region}(X) = \text{"Chhattisgarh"} \Rightarrow$$

$$(\text{Dimension}(X) = \text{"Impressions"})[t : 0.29\%]$$

$$\vee (\text{Dimension}(X) = \text{"Clicks"})[t : 9.42\%]$$

(2)

$$\forall X, \text{Region}(X) = \text{"Tamilnadu"} \Rightarrow$$

$$(\text{Dimension}(X) = \text{"Impressions"})[t : 48.88\%]$$

$$\vee (\text{Dimension}(X) = \text{"Clicks"})[t : 42.58\%]$$

(3)

Above discrimination rules indicates weightage of impressions and clicks in total clicks and impressions of selected regions. It indicates main two regions of the country India. Chhattisgarh is the region where the usage of the website is very low and Tamilnadu is the region where the usage of the website is highest among all the regions.

Region wise Analysis of Clicks and Impressions is shown in Figure 1 clearly indicates that there are so many customers opening the website and very few of the customers are clicking on the pages of the website. The difference shows that in the states where the website is used very less they are definitely clicking on the web pages of the website.

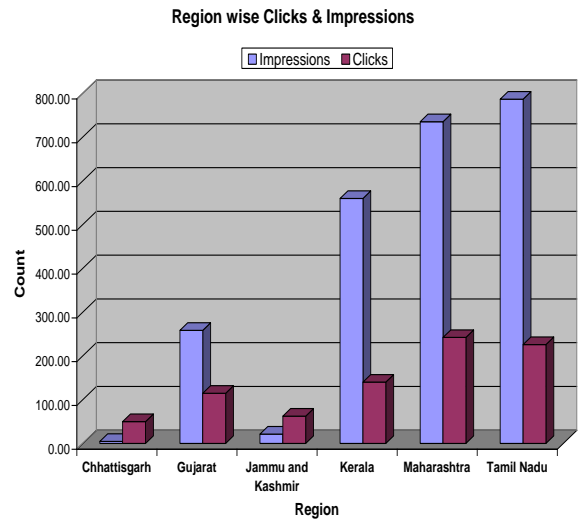


Figure 1. Region wise difference between Impressions and Clicks

Customer behavior can be analyzed using the graph given in Figure 2. It shows the percentage of the users and their interest in the products of the company. We can see that Tamilnadu is the state where usage is higher and impressions are very high. But the clicks on the web page are lower in comparison to other states impressions.

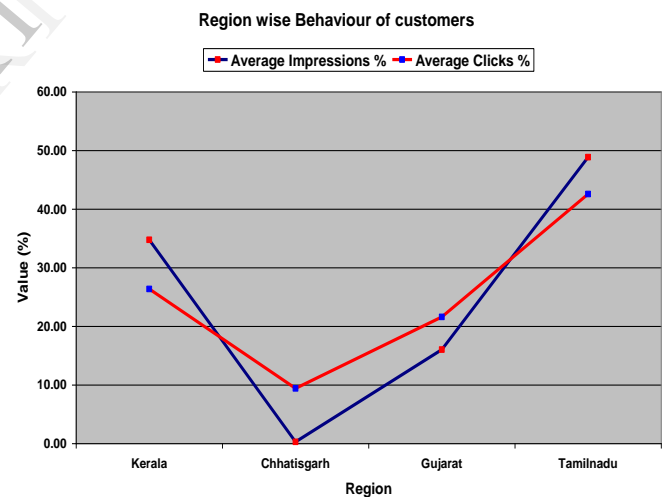


Figure 2. Regionwise Customer Behaviour

VI. RESULT ANALYSIS

Analysis of Impressions and Clicks shows that the users open the website but every time they do not click on the pages of the website. It means the company needs to restructure the website for personalization and make it more users friendly. On the basis of usage in the States of India, the owner of the company can plan future advertisement campaigns and increase customer base in those states. The difference between the Clicks and Impressions shows that in some States where the website is opened rarely, their click ratio higher than other states where impressions are higher.

VII. CONCLUSION

The study of the Impressions and Clicks shows that the ratio of Clicks in comparison to Impression is very low. But it can be increased by better layout and attractive animation on the website. The cost spent for the campaign is negligible in comparison to the benefit of the company. The company can easily get idea about their potential customers in the country and also helps to focus specified region in the country.

REFERENCES

- [1] Ankit R Kharwar¹ and Viral Kapadia, "A Effective Preprocessing for Web Usage Mining", Department of Computer Engineering, Birla Vishvakarma Mahavidyalaya College of Engineering Anand, Gujarat, India.
- [2] Zhenhuan He, "The Study of Personalized Recommendation Based on Web Data Mining", School of Information Engineering Nanchang Hang Kong University Nanchang, China.
- [3] Mahendra Pratap Singh Dohare¹, Premnarayan Arya², Aruna Bajpai³, "Novel Web Usage Mining for Web Mining Techniques", Department of Software System, Samrat Ashok Technological Institute Vidisha, M.P., India.
- [4] Neil Daswani, Chris Mysen, Vinay Rao, Stephen Weis, Kourosh Gharachorloo, Shuman Ghosemajumder, and the Google Ad Traffic Quality Team "Online Advertising Fraud" From the forthcoming book, Crimeware, edited by Markus Jakobsson and Zulfikar Ramzan, 2008
- [5] Mobasher B., Dai H., Luo T, Nakagawa M. Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization. Data Mining and Knowledge Discovery, 2002
- [6] Qin Fengrui, "Personalized recommendation engineering research and in Digital library application ," Changchun: Changchun University of Science and Technology, 2010 .
- [7] Lu Lina and Yang Yi Ling, "In Web diary minning data pretreatment research," Computer project, 2000
- [8] Mobasher B et al., "Integrating web usage and content mining for more effective personalization," Proceeding of the E C-WEB Conference. Springer, 2000

IJERT