International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 8 (2015) pp. 20817-20822

http://www.ripublication.com

Analyzing Consumer Behavior Towards E-Commerce Website Using Web Usage Mining

T. Archana and Dr. Ajay Kumar Sharma

Computer Applications Division VIT University Vellore, India
Commerce Division, VIT University, Vellore -14.
archana.t@vit.ac.in

Abstract

The continuous growth of electronic commerce websites, its success and failure has stimulated great interest in analyzing consumer behavior towards e-commerce websites. Given the significant rise in the online users, better understanding of consumers allows better marketing strategies to be designed. Conventional methods are out dated in accessing the consumer behavior towards e-commerce websites. In this paper web usage mining is used to extract the consumer behavior from the web log file and K-Means Clustering algorithm is used to analyze the consumer behavior data.

Keywords: Web usage mining; online consumer behavior; e-commerce; log files

Introduction

The web continues to grow at an incredible rate. As a result traditional "brick and mortar" stores are stepping into e-commerce to serve a variety of consumers covering a wider geographical area. Among the factors which decides the success, survival and failure of this e-commerce website is consumer expectation and satisfaction. A better understanding of the consumer paves way for a better marketing strategy to win the market. As a consequence, a large number of studies on discovering new models, theories and knowledge on internet consumer behavior have gained focus.

With wide spreading data on online consumer behavior data mining techniques are employed to extract information and knowledge which is hidden in a large incomplete and vague dataset. Web mining is one application of data mining technology, which extracts interesting and potentially useful patterns and hidden information from web documents and web activities [1]. Web Usage Mining is a part of web mining which deals with the extraction of interesting knowledge from log files produced by web

server. Web Usage Mining is otherwise termed as Web Log Mining or Web Usage Analysis or Click Stream Analysis.

In this paper web server log file is taken as base data and web usage mining tool is employed to extract the required user profile information from this log file. To cluster the user profile information into segment K-means algorithm is applied to get a distinct cluster segment.

Literature Review

Zekulin states "Data mining is the process of extracting previously unknown , comprehensive and actionable information from large database and using it to make crucial business decisions"

Ferruzza states "Data mining is the set of methods used in the knowledge discovery process to distinguish previously unknown relationships and patterns within data"

A new concept of mining called "Web mining" is proposed by Etzioni in 1996, using data mining techniques he extracted information from vast data on the WWW.

Srivastava et al [1] termed Web usage mining as a task of applying data mining technique to discover usage patterns from web data in order to understand and better serve the needs of the users navigating on the web.

Martin –Guerrero et al [2] conducted a comparative study on the performance of different clustering algorithms such as K-Means, Hierarchical clustering, fuzzy C-Means and SOM on websites. The result of the simulation suggested that K-Means algorithm was adequate for a simple website and SOM will be preferred for a complex website.

Yang et al [3] suggested K-Means algorithm may be treated best in clustering user sessions for web catching and pre fetching.

Web Mining Model

Web mining generally has a wide range of applications in discovering and extracting the hidden information from data stored in the web. Web mining is used to discover the user information about the user activity stored in various format of the web.

Web mining can be categorized into three classes based on the mode of mining done on the web. They are 1) Web Usage mining 2) Web Content Mining 3) Web Structure Mining.[4,5,6]

Web Usage Mining [1] is the process of mining the activities of the user while they are browsing or navigating through the web. Web Usage Mining helps the e-commerce websites to enhance the quality of the website design and its services and to coin new marketing strategy.

Web Content Mining [7,8]is an efficient mechanism to help the user to mine the information from the contents of the web. It also organizes and clusters the document on the web and hence helping the search engine in accessing the documents by keywords.

Web Structure Mining [9,10] is the process of mining the structure of the hyperlink within the web. It helps in information retrieval and improves classification of documents.

A. Data preprocessing

The actual data collected from the web log file contains the browsing history of the user browsing behavior. This actual data cannot be directly fed as input for Mining algorithm because it consist of redundant and inappropriate data.

To ensure the accuracy and efficiency of the information required for decision making, the quality of log file data is improvised by applying the data preprocessing technique.

Data preprocessing includes Data Cleaning, User Identification and user session identification.

- A serious problem caused by redundant data is Anomalies. Detecting and removing the anomalies and reducing the redundancy is done through data cleaning
- In order to perform clustering each user should be distinctly identified. This is achieved through user registration technique.
- Based on the user identification , user access information is further divided into distinct session processes.

B. Data Analysis

Mine Set by Silicon Graphics has perhaps the best balance between powerful knowledge discovery functions and an impressive 3D interactive visualization [11] for more than a dozen mining tasks.

Mineset tool is used for data analysis. The dataset is parted into two on the ratio 7:3 and utilized as training data and testing data respectively.

Data is segmented on the basis of four factors –age, gender, online transaction time and online transaction type.

Age was segmented into 4 segments as

1. Age between 15 to 20
2. Age between 20 to 25
3. Age between 25 to 30
4. Age between 30 to 35
5. Age between 35 to 40

Gender was segmented into 2 segments

1. Male
2. Female

Online transaction time was segmented into 4 segments

1. 12.00 am to 5.59 am
2. 6.00 am to 11.59 am
3. 12.00 pm to 5.59 pm

4. 6.00 pm to 11.59 pm

Online transaction type was segmented into 4 segments

1. Electronic items includes Purchasing Mobile, Computer, Laptop, tablets, gadgets etc
2. Clothes includes wearable clothes of all kind
3. Accessories includes leather wearable items, metal wearable items ,wallets etc.
4. Others includes house hold, toys, gifts etc.

Table 1: Clustering Customer Behavior

TYPE	**CLUSTER**			
	1	***2***	***3***	***4***
AGE				
15-20	0	2	3	0
20-25	8	35	93	4
25-30	23	33	87	8
30-35	63	31	20	28
35-40	57	8	13	25
GENDER				
MALE	98	38	115	50
FEMALE	53	70	101	15
TRANSACTION HOURS				
0.00-5.59	0	16	6	0
6.00-11.59	61	33	26	9
12.00-17.59	68	38	98	30
18.00-23.59	23	22	86	26
TRANSACTION TYPE				
ELECTRONICS ITEM	106	10	108	53
CLOTHES	18	35	10	10
ACCESSORIES	24	15	91	2
OTHERS	3	30	7	0

C. Result analysis

Cluster 1:

This cluster is formed by Male consumers whose age ranges between 30 to 40 and they purchase Electronic items during 6.00 to 18.00 hours

Cluster 2:
This cluster is formed by Female consumers whose age lies in the range 20 to 35 and who purchases clothes and others during 6.00 to 18.00 hours

Cluster 3:
This is the largest cluster among the rest formed by Male and Female consumers of age group 20 to 30 who shows interest in purchasing Electronics and Accessories during 12.00 to 24.00 hours

Cluster 4:
The smallest cluster is formed by Male consumer of age 15-20 who buy Electronics item during 18.00 -24.00 hours.

Conclusion

E-commerce companies are moving into a new world of variety and customized product for each consumer. People started preferring online store to satisfy their personalized demands from a wide range of products available across the globe. Hence to personalize the product for the consumer e-commerce company has to go for a personalized consumer target marketing, which may be achieved only by studying and analyzing the behavior of the consumer.

It is found in this study that the majority of the consumers who is involved in e-commerce were Male and their browsing hour's ranges from 6.00 am to 6.00 pm. Electronic items and accessories are purchased by both Male and Female whereas the clothes are purchased in large number by Female consumers whose age ranges from 20 to 30. Online purchase of House hold and other items are less and hence the ecommerce company has to find suitable marketing strategy to target this gap found in selling household and other items.

REFERENCES

[1] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data," SIGKDD Explorations, Vol. 1, No. 2, pp. 12-23, 2000

[2] J.D. Martin-Guerrero et al, Studying the feasibility of a recommender in a citizen Web portal based on user modeling and clustering algorithm, Expert Systems with Applications, 30 (2) (2006), pp. 299–312

[3] Q. Yang, J.Z. Huang, M. Ng, A data cube model for prediction-based Web prefetching, Journal of Intelligent Information Systems, 20 (1) (2003), pp. 11–30

[4] Kosala and Blockeel, "Web mining research: A survey," SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining, ACM, Vol. 2, 2000

[5] K. Madria, S. S. Bhowmick, W. K. Ng, and E.-P. Lim, "Research issues in web data mining," in Data Warehousing and Knowledge Discovery, 1999, pp. 303-312

[6] J. Borges and M. Levene, "Data mining of user navigation patterns," in WEBKDD, 1999, pp. 92-111

[7] S. Chakrabarti, "Data mining for hypertext: A tutorial survey." SIGKDD: SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining, ACM, Vol. 1, No. 2, pp. 1-11, 2000

[8] M. Balabanovic and Y. Shoham, "Learning information retrieval agents: Experiments with automated web browsing," in Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogenous, Distributed Resources, 1995, pp. 13-18

[9] J. Hou and Y. Zhang, "Effectively finding relevant web pages from linkage information." IEEE Trans. Knowl. Data Eng., Vol. 15, No. 4, pp. 940-951,2003

[10] H. Han and R. Elmasri, "Learning rules for conceptual structure on the web," J. Intell. Inf. Syst., Vol. 22, No. 3, pp. 237-256, 2004

[11] Barry G. Becker. Visualizing Decision Table Classifiers.In Graham Wills and John Dills, editors,Proceedings of Information Visualization '98,pages 102-105, Los Alamitos, CA, Oct 19 – 20,1998. IEEE CS Press.