SCITECH
RESEARCH ORGANISATION|
Volume 2, Issue 1
February, 2015|
Journal of Information Sciences and Computing Technologies
www.scitecresearch.com

# A Study of Customer Behaviour Through Web Mining

Gan Teck Wei[1], Shirly Kho[2], Wahidah Husain[3], Zurinahni Zainol[4]
[1,2,3,4] School of Computer Sciences, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
[1]gtwei.ucom11@student.usm.my, [2]skho.ucom11@student.usm.my, [3]wahidah@cs.usm.my, [4]zuri@cs.usm.my

## Abstract

Web mining is the extraction of interesting and potentially useful patterns and hidden information from web documents and web activities by applying data mining technology. The most important challenge of electronic commerce (E-commerce) is to understand as much as possible the customers' wants, desires, and buying patterns to ensure competitiveness in the E-commerce era. Nowadays, any information related to consumer behavior has an important value in the highly competitive nature of the E-commerce market. Therefore, web mining can be used to find those obvious data that have potential value to reduce competition and simultaneously increase business profit. This paper aims to study the classification of web mining to extract customer behavior in E-commerce, investigate customer behavior through the techniques and processes of the web data mining used, explore the application of web mining in E-commerce, and increase profit.

**Keywords** Data mining; Web mining; Buyer pattern; E-Commerce; Customer pattern

## 1. Introduction

Given the increasing popularity of the Internet and the rapid development of E-commerce, Internet-based businesses' websites are facing increasing competition. E-commerce sites generate large amounts of data daily, and these data include potential consumer-related information that is valuable for market analysis and prediction [1]. Web data can be web server access logs, proxy server logs, browser logs, user registration information, payment transactions, and users' sessions [2]. Therefore, the most important challenge of E-commerce is to elucidate customers' wants, love, and value orientation as much as possible to ensure competitiveness in the E-commerce era.

Web mining is the extraction of interesting and potentially useful patterns and hidden information from web documents and web activities by applying data mining technology [3]. Nowadays, any information related to consumer behavior has an important value to the highly competitive nature of the E-commerce market. Therefore, web mining can be used to find those obvious data that have potential value to reduce competition and simultaneously increase business profit [4].

In the field of science and technology management, web mining has a more important value and meaning: extracting and using rich information resources from the Internet is the key to successful technological workers and decision makers [5]. The application of data mining or other information process techniques to the World Wide Web (www) through useful patterns of major customers reduces competition and simultaneously increases business profit in E-commerce [6].

## 2. Classification of Web Mining

Web data mining is applied by using data mining technology in a web environment. It is basically classified into three categories, namely, web content mining, web structure mining, and web usage mining, as shown in Figure 1 below.

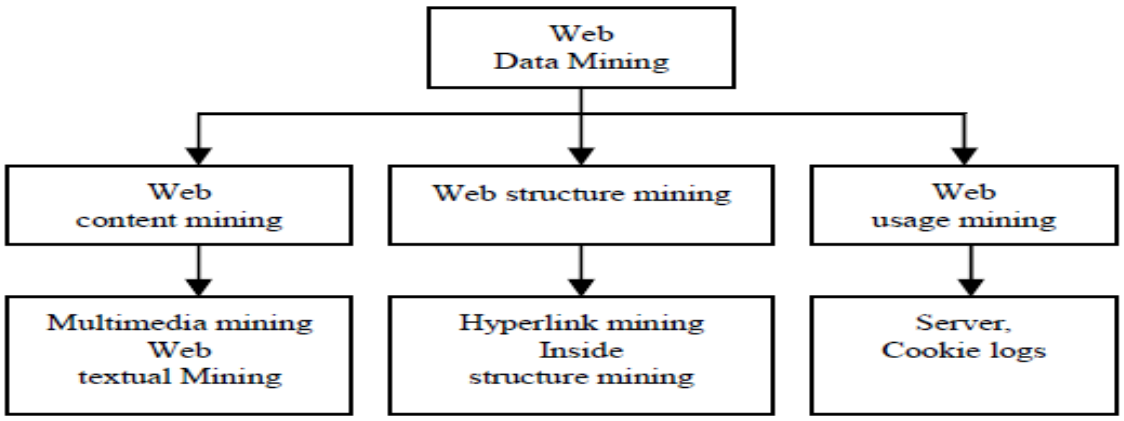


**Fig 1: Web Mining Classifications [7]**

### 2.1 Web Content Mining

Web content mining is the process of finding and extracting information from a vast amount of collected web data. This process can be classified into text mining and multimedia mining. Web content mining involves re-processing to obtain more accurate and useful information query results from a search engine. The semi-structured web information will be reconstructed to a more structured data, and then the standard database query mechanism and the method of mining can be used for analysis. A summary of page, classification, clustering, and association rules can be found when mining HTML page content using text mining to the text on the page or using multimedia data mining to the multimedia information on the page.

### 2.2 Web Structure Mining

Web structure mining basically finds potential web link structure model. It can work by deriving knowledge from the organizational and link relationship on the www, analyzing the webpage link visited by the user, analyzing the number of links and link objects, and creating their own web link structural mode. This model can be used for web page classification, and then the similarity and the relationship to access the information between different pages can be found to investigate the rules of web pages related to the subject.

### 2.3 Web Usage Mining

Customers' business activities and their browsing information on a website are recorded in a log file. Web usage mining analyzes the web page browsing model of customers and finds the information of potential customers according to the server's log files, which are used to understand and handle customers' web behavior.

The web site structure can also be improved once customer behavior is understood. Two types of tracking area exist: general access pattern tracking and personal usage records tracking. General access pattern tracking is used for improving the website organizational structure on understanding customers' access patterns and trends. Personal usage records tracking analyzes the preferences of individual customers and provides customized websites for different customers, depending on their access patterns [8].

## 3. Techniques and Process of the Web Data Mining

A large data should pass through four stages to analyze each customer action within the web system as shown in Figure 2. Given the reliable information, the company is able to enhance its business value and quality based on the information on customer behavior. The four stages of web mining are as follows: source data collection, data preprocessing, pattern discovery, and pattern analysis.

### 3.1 Source data collection

Raw data are mostly collected from customers who use the web system, and the customers' data will be stored in the web log files on the web server. Therefore, the web server is the main source of data for web mining. Web log files basically record visitors' browsing behavior in the history [4] and collect raw data from the client side, the server side, and the agent in the web log.

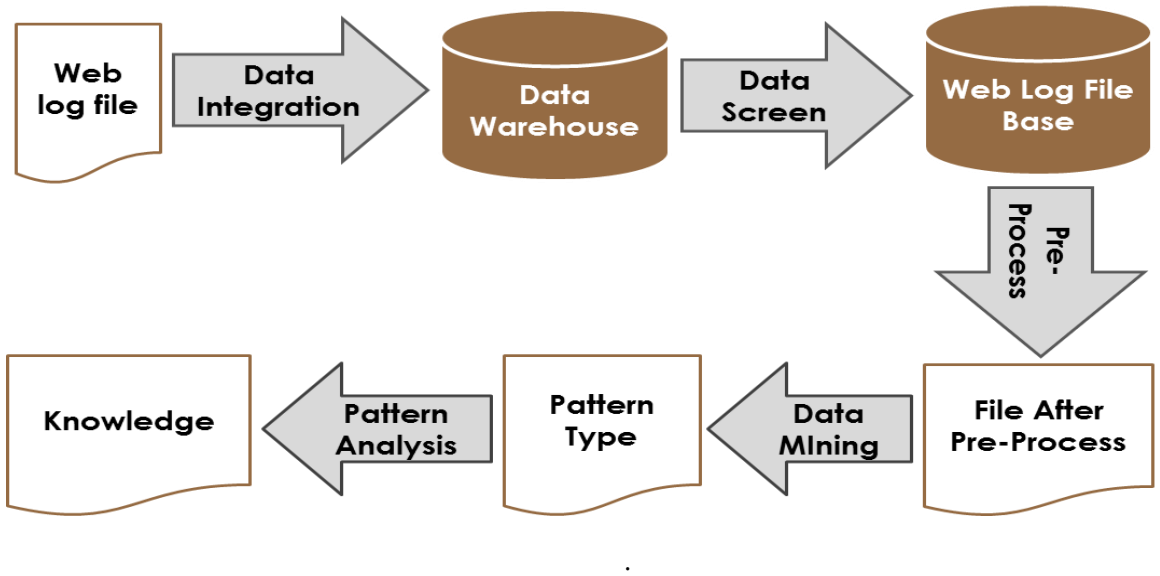


**Fig 2: Process of the web mining**

### 3.2 Data Pre-processing

Some of the features that may be included in the actual data collected are ambiguity, redundancy, and incompleteness [9]. Preprocessing is very important in making the further processes more effective and better. Preprocessing of large collected data is necessary. The preprocessing includes data cleaning, user identification, user session identification, access path supplement, and transaction identification. Data cleaning mainly aims to remove redundant data from the web log to reduce the scope of the objects. After data cleaning, user identification is performed to identify the users individually. The techniques that will be used are user registration techniques, cookies that store user information, and heuristic rules. User session identification aims to separate users' access information into numerous sessions using time-out estimation. Access path supplement should be added in the preprocessing phase to collect the user access patterns accurately. To obtain adequacy for the desired results of the web mining analysis, transaction identification is used to divide or oppositely combine user transactions.

### 3.3 Pattern Discovery

This stage aims to find the hidden relationship in the data. Pattern discovery basically includes four techniques: classification analysis, association rule discovery, sequential pattern discovery, and clustering analysis. Classification analysis is classified according to the preprocessed data to categorize the profiles of the customers. Personalized information services and activities will start after classification. Association rule discovery in web mining is used to discover the most frequently requested pages in user sessions. The design of the website can be improved by using mining information based on the group of data on the pages that users visit [10]. Sequential pattern discovery is the an important process in web mining. It can be used to forecast customers' behavior to mine the data from web logs that have records of the customers' visit patterns. The company can provide personalized service for a customer to increase the interaction between the company and the customer and also to retain existing customers. Clustering analysis technique is the most popular technique used in web mining applications. This technique groups the data, which contain similar features and characteristics, to facilitate marketers in their marketing decisions. Two clustering methods are used: partition clustering and hierarchical clustering.

### 3.4 Pattern Analysis

In pattern analysis a set of patterns is generated by model pattern discovery algorithm. This algorithm selects an interesting pattern to find other valuable models. Other patterns will be ignored.

## 4 Application of Web Mining in E-Commerce

In the E-commerce market, any mined information of customer behavior is valuable to businesses. Based on the information, the customers' most frequently accessed pages, preferences, interests, and value orientations are known by the market. Therefore, a merchant area can use the information to adjust the structure of its website, build the image of the company, improve the service to attract new customers, and build an appropriate E-commerce platform to satisfy the visitors' demands.

### 4.1 Customer Analyzing

Mined data help acquire new customers that will become future potential customers of merchants by analyzing the web log that defines whether a new customer is profitable. In addition to retaining existing customers, predicting customer purchase behavior allows improvement of merchant services and profit increase [11].

### 4.2 Website optimization

This step is important in E-commerce; when customers use their browser to go through business' sites, the websites will provide dynamic advice, such as product recommendations, to be adopted by customers' sites. This technique will allow customers to have different feelings when they enter a website. To provide personalized services, website optimization uses access behavior modeling and user access behavior forecasting to obtain customers' spending habits, likes, and other characteristics.

### 4.3 Web Personalization

According to the information from user behavior, a website can be designed and re-structured to make it more advance and user-friendly. In addition, the image and product value of the company is very important in satisfying customer need based on website quality. Therefore, a merchant can improve its service quality and performance to satisfy its customers by using the behavior information [10]. The obtained data will help improve the performance of a website, smoothen network communication, and accelerate the response of the system.

### 4.4 Business Intelligent

In web technology, a technique called customer visit pattern is becoming very important and has the potential for E-commerce marketing. It helps the merchant to know well the customer purchase objective in their website and their customers' consumption demand.

## 5 Discussion

The application of web mining in E-commerce has many benefits. The business environment keeps on moving forward, and customer's cogitation are changed, especially in terms of different methods of purchasing. Some companies still operate in the same place and refuse to accept new technology. To increase profitability and enhance the service provided to the customer, understanding customer behavior is crucial for companies who move forward to the future business opportunities. To date, customers are more concerned about the service provided by companies and hope that companies know what they need. According to the mining results, companies can improve their website services and productivity, increases their potential customers, and effectively knows the action of the customers when they browse their websites [12]. Adopting web mining in E-commerce requires much knowledge on the latest technology and a good understanding on how to use mining techniques and algorithms in every step of the web mining process. Good understanding of the application skills and employees' experiences determine the success in identifying customer behavior well and also the future of the company.

## 6 Conclusion

Web mining is a technology that has strong practical significance in E-commerce. This technology not only helps enhance the performance of the website and understand customer need, but also serves as a basis for enhancing the topology of the site and hyperlinks. This research discusses the classification of web mining, which is classified into three categories: web content mining, web structural mining, and web usage mining. The web mining process in terms of how to mine a large number of data to obtain customer behavior information is also discussed. It also discusses the importance of the application of web mining to E-commerce, which has a major influence to the merchant, customer, and company.

## References

[1] Li Yong-hong; Liu Xiao-liang, "Research of data mining based on e-commerce," Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on , vol.4, no., pp.719,722, 9-11 July 2010

[2] Chaoyang Xiang; Shenghui He; Lei Chen, "A Studying System Based on Web Mining," Intelligent Ubiquitous Computing and Education, 2009 International Symposium on , vol., no., pp.433,435, 15-16 May 2009

[3] Li Mei; Feng Cheng, "Overview of Web mining technology and its application in e-commerce," Computer Engineering and Technology (ICCET), 2010 2nd International Conference on , vol.7, no., pp.V7-277,V7-280,16-18 April 2010

[4] Yadav, M.P.; Feeroz, M.; Yadav, V.K., "Mining the customer behavior using web usage mining in e-commerce," Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on , vol., no., pp.1,5, 26-28 July 2012 Dsfds

[5] Gu Chengjian; Huang Lucheng, "Web Mining in Technology Management," Business and Information Management, 2008. ISBIM '08. International Seminar on , vol.2, no., pp.88,91, 19-19 Dec. 2008

[6] Li Haigang; Yin Wanling, "Study of Application of Web Mining Techniques in E-Business," Service Systems and Service Management, 2006 International Conference on , vol.2, no., pp.1587,1592, 25-27 Oct. 2006

[7] Cheng Yu; Xiong Ying, "Application of Data Mining Technology in E-Commerce," Computer Science-Technology and Applications, 2009. IFCSTA '09. International Forum on , vol.1, no., pp.291,293, 25-27 Dec. 2009 Ff

[8] Zhiwu Liu; Li Wang, "Study of Data Mining Technology Used for E-commerce," Intelligent Networks and Intelligent Systems (ICINIS), 2010 3rd International Conference on , vol., no., pp.509,512, 1-3 Nov. 2010

[9] S.Yadav, K.Ahmad, J.Shekhar, ”Analysis of Web Mining Applications and Beneficial Areas”, International Journal of IUM, vol.12, Issue 2, 2011.

[10] Shen Zihao; Wang Hui, "Research on E-Commerce Application Based on Web Mining," Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on , vol., no., pp.337,340, 22-23 June 2010 Fsd

[11] Tang, Hewen; Yan, Honglin; Zengfang Yang; Yu Ma; Chunping Li, "Application of Data Mining in Electronic Commerce," Information Engineering, 2009. ICIE '09. WASE International Conference on , vol.1, no., pp.631,634, 10-11 July 2009

[12] Preeti Sharma and Sanjay Kumar, “An Approach for Customer Behavior Analysis Using Web Mining”, International Journal of Internet Computing, ISSN: 2231-696, Issue-2, 2011.