Intelligent E-Business to Increase Online Sales Using Web Mining

Sachin Balvir[1], Santosh Sahu[2], Jaynat Rohankar[3]

[1]*Assistant Professor, Department of Information Technology, DMIETR, Wardha, India*

[2]*Research Scholar, Department of Computer Science Engineering, VNIT, Nagpur, India*

[3]*Assistant Professor, Department of Computer Science Engineering, TGPCET, Nagpur, India*

ABSTRACT

Data mining software allows users to analyze large databases to solve business decision problems. Data mining is, in some ways, an extension of statistics, with a few artificial intelligence and machine learning twists thrown in. Ever rising competition and changing demands of customer it is difficult to satisfy the people with different backgrounds. E- Marketing with its various advantages of low cost, high efficiency, time saving has assumed importance these days. Like statistics, data mining is not a business solution; it is just a technology. This paper provides the concise outline how the user access behavior is used increase sales.

Keyword : - *Web mining, Online sales, e-business.*

1. INTRODUCTION

The wide use of the Internet has essentially changed the ways in which we communicate, collect information and make purchases. As the utilization of the World Wide Web (WWW) and exchange of email, data over internet increased radically, the computer scientists hurried to describe this new phenomenon. In the beginning they were shocked by the incredible mixture the Internet confirmed in the size of its features, they soon exposed a general pattern in their capacity: there are a lot of small elements enclosed within the Web, but only some large ones. A few sites have of millions of pages, but millions of sites only have a handful of pages. Few sites have millions of links, but many sites contain one or two. Millions of users gather to a few select sites, giving little awareness to millions of others.[1]With the volatile increase of information sources offered on the World Wide Web, it has become more and more essential for users to utilize computerized tools in order to find, extract, filter, and evaluate the preferred information and resources. In addition, with the revolution of the Web into the main tool for electronic commerce, it is essential for organizations and companies, who have invested millions in Internet and intranet technologies, to follow and evaluate user behavior. These factors give rise to the need of creating client-side and server-side intelligent systems that can successfully extract knowledge. Many organizations make available information and services on the web such as on-line shopping, customer support, web based applications etc. are becoming common practice. The WWW is becoming everywhere and a regular tool for daily activities of common people, from a child to a senior across the world [2].

The business benefits that web mining affords to digital service providers include personalization, collaborative filtering, enhanced customer support, product and service strategy definition and fraud detection. Today business talks are more and more on e-business as they incorporate internet technology into their core business processes. This new and modern business requires the key web mining process to be merged with the new technologies [11].

2. VARIOUS SOURCE OF DATA

Web Mining is based on knowledge discovery (KD) from web. It extracts the knowledge & represents in a proper way. Web mining is like a graph & all pages are node & each connects with hyperlinks. Web mining is useful to extract the information, image, text, documents etc. Sources of data in e-business are shown in fig 1 and brief explanation as follows [9]:

a. *Web Pages:*
Web pages usually include text, links, images, graphics, audio & video. Web pages are the main source of data for e-business.

b. *User Registration:*
When user wants to create account then he or she has to register itself. Here user has to fill different types of information. It usually contains demographic information and personal interests and preference of the person.

c. *Web Server Log:*
Website statistics are based on server logs. A server log is a simple text file which records activity on the server. There are several types of server log — website owners are especially interested in access logs which record hits and related information. Information about the request, including client IP address, request date/time, page requested, HTTP code, bytes served, user agent, and referrer are typically added.

d. *Event Log:*
Event log is also an important data source. It logs the information such as advertisements seen, products added to shopping cart, products bought, etc.

e. *Customer Queries:*
These are the search terms used by the customer. This data is produced on e-business web servers. These types of data can be very useful.

f. *Transaction Database:*
It includes details such as customer id, products ordered, time, quantity, price etc,

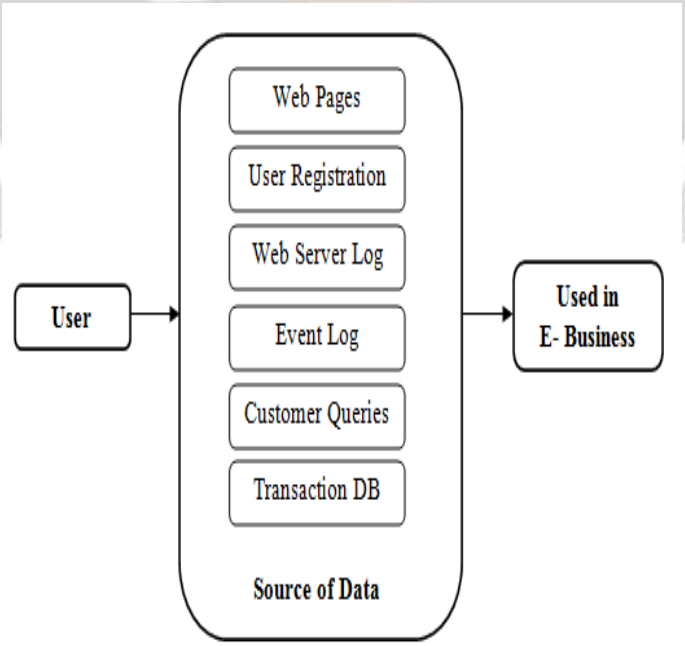

Fig. 1 Sources of data used in E-Business

3. OVERVIEW OF WEB MINING

A. Function of Web Mining:

To clarify the uncertainty to find out what forms Web mining. Kosala and Blockeel [4] had recommended a decomposition of Web mining in the following tasks :

1. *Resource finding*: the task of retrieving intended Web documents.
2. *Information selection and pre-processing*: automatically selecting and pre-processing specific information from retrieved Web resources.
3. *Generalization*: automatically discovers general patterns at individual Web sites as well as across multiple sites.
4. *Analysis*: validation and/or interpretation of the mined patterns.

B. Web Mining Classifications:

Extraction of interesting information or patterns from large databases is called Data Mining. Web Mining [3] is the application of data mining techniques to discover and retrieve useful information (knowledge) from the WWW documents and services. Web mining can be divided into three categories [5] namely web usage mining, web content mining and web structure mining.

a) Web Structure Mining (WSM):

Web structure mining, one of three categories of web mining for data, is a tool used to identify the relationship between Web pages linked by information or direct link connection. This connection allows a search engine to pull data relating to a search query directly to the linking Web page from the Web site the content rests upon. This completion takes place through use of spiders scanning the Web sites, retrieving the home page, then, linking the information through reference links to bring forth the specific page containing the desired information. [7].

b) Web Content Mining (WCM):

Web content mining, also known as text mining, is generally the second step in Web data mining. Content mining is the scanning and mining of text, pictures and graphs of a Web page to determine the relevance of the content to the search query some ranking mechanism (web mining) either This scanning is completed after the clustering of web pages through structure mining and provides the results based upon the level of relevance to the suggested query. With the massive amount of information that is available on the World Wide Web, content mining provides the results lists to search engines in order of highest relevance to the keywords in the query. The main uses for this type of data mining are to gather, categorize, organize and provide the best possible information available on the WWW to the user requesting the information.

c) Web Usage Mining (WUM):

Web usage mining is the third category in web mining Web usage mining is used to discover user navigation patterns and the useful information from the web data present in server logs, which are maintained during the interaction of the users while surfing on the web. Most existing Web analysis tools provide mechanisms for reporting user activity in the servers and various forms of data filtering. Using such tools it is possible to determine the number of accesses to their server and to individual files, the times of visit and the domain names and u rl‟s of users.

C. Methods of Web Mining

There are different types of techniques are offered for data mining. The most frequently used techniques are artificial neural networks, decision trees, and the nearest-neighbor method visualization, association rule, classification and clustering. each one of these approaches brings special advantages and disadvantages that have to

be considered prior to their use [6]. A good way to apply advanced data mining techniques is to have a flexible and interactive data mining tool that is completely incorporated with a database or data warehouse..In web mining some of the techniques of data mining can be used for example association rule, classification and clustering etc. The brief descriptions of these techniques are as follows.

Classification:

Classification is use to build up a idea of the type of customer, item, or object by relating several attributes to categorize a particular class. For example, one can easily classify cars into different types (sedan, 4x4, convertible) by identifying different attributes (number of seats, car shape, driven wheels). One can apply the same principles to customers, for example by classifying them by age and social group. Classification algorithms can be used to categorize users into special classes according to their browsing behavior or pattern. The criterion by which items are assigned to different clusters is the degree of similarity among them.

Prediction:

Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other web mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event.

Sequential Pattern:

Sequential patterns are a helpful technique for identifying trends, or regular occurrences of similar events. For example, with customer data we can identify that customers buy a particular collection of products together at different times of the year. In web usage mining, sequential patterns are exploited to find sequential navigation patterns that appear in users sessions frequently.

Decision Tree:

The objective is to build a model that predicts the value of a target variable based on several input variables. A decision tree is a simple representation for classifying examples. Decision tree learning is one of the most successful techniques for supervised classification learning

Association Rules:

The association rule method can be used to show pages that are mainly referenced together and to discover the direct or indirect relationships between web pages in users browsing behavior. After transactions are detected in the preprocessing phase, frequent item-sets are discovered using the each hyperedge is weighted by the averaged confidence of all the possible association rules formed on the basis of the frequent item-set that the hyperedge represents [2].

4. INTELLENGENT E-BUSINESS PROCESS

The drastic development of internet and information technology made E-business which is a new type of commercial channels developing prosperously [8]. How to analyze data of E-business users for mining users' information that enterprises interest is critical for their development. Web mining has advantages that it can mine data efficiently and intelligently, so it is becoming more and more important in modern E-business.

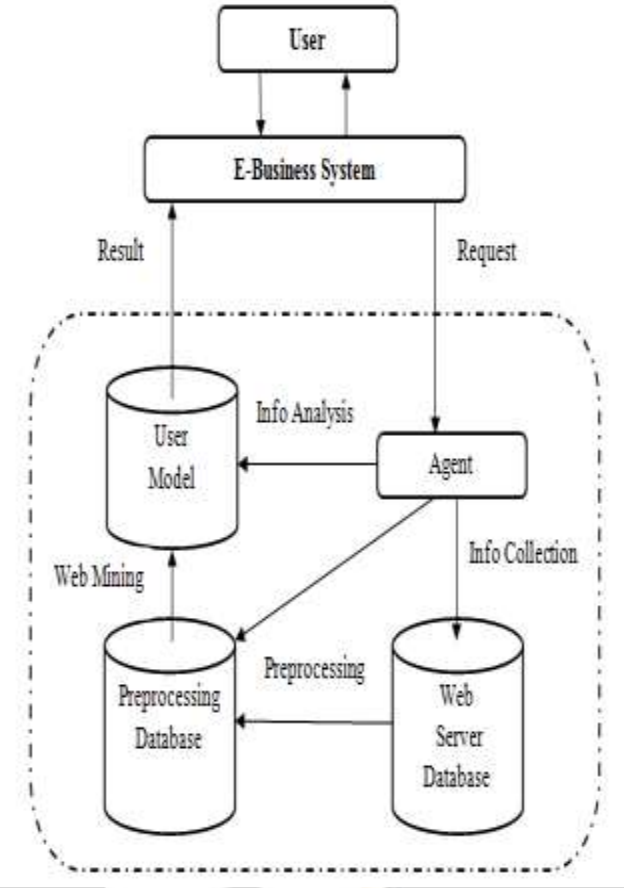

Fig. 2 Architecture of E-Business

Fig 2 shows the architecture of e-business. Work flow of e-business architecture is as follows [10]:

i. A user logs on to an e-business website platform. Then, user Agent enables this function module on the basis of user usage information.
ii. According to the personalized requirements of users, organize data resource and find the original use data of this user.
iii. Preprocess the user data, including data cleansing, conversion, integration and formatting, and load the results to the preprocessing data resource bank.
iv. Select a suitable data mining method in collaboration with user Agent to build user model and model base.
v. Based on data mining results, integrate with expertise and area rules, and offer users personalized e-business services via an e-business system.

5. CONCLUSION

It is easy to collect data from web-enabled e-business sources as all visitors to a web site leave the trail which automatically is stored in log files by web server. The data mining tools can process and analyze such web server log files or actual web contents to discover meaningful information. E-business with the help of internet grows quickly; personalized e-business should be worth paying more concentration and increasing from the theoretical and practical standpoints. It is easy to collect data from web-enabled e-business sources as all visitors to a web site leave the trail which automatically is stored in log files by web server. The data mining tools can process and analyse such web server log files or actual web contents to discover meaningful information. E -Business holds an important key to

every organizations future. This paper presents an architectural framework for personalized e-business service recommendation system using data mining techniques.

6. REFERENCES

[1] Kavita Sharma, Gulshan Shrivastva and Vikas Kumar, "Web Mining: Today and Tomorrow", 3rd International Conference on Electronics Computer Technology (ICECT 2011), In IEEE, 2011.

[2] Manoj Padia, Shubhendhu Kumar Pani, Sanjay Kumar Padhi, Lingaraj Panigrahy and R. Ramakrishna, "A Review of Trends in Research on Web Mining", International Journal of Instrumentation , Control & Automation (IJICA), volume 1, Issue 1, 2011.

[3] Raymond kosala, Hedrik Blockeel, "Web Mining Research: A Survey", Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium.

[4] Raymond Kosala and Hendrik Blockeel, "Web Mining Research: A Survey", ACM SIGKDD Explorations Newsletter, June 2000, Volume 2, Issue 1.

[5] R.Cooley, B.Mobasher and J.Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web". In Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), 1997.

[6] Mr. Dushyant B. Rathod and Dr. Samrat Khanna, "A Review on Emerging Trends of Web Mining and Its application", International Journal of Engineering Development and Research (IJEDR).

[7] S. U. Balvir, G. N. Tikhe and L. M. Barapatre, "Positioning Webpage Using Rank", International Journal of Computer Technology and Electronics Engineering (IJCTEE), 2011.

[8] Ana Maria Mihaela Tudorache, Romanian-American University "Data Mining and the process of taking decisions in e-business"

[9] Pankaj Gupta and Bharat Bhushan, "Data Mining Techniques in e-business", International Journal of Advances in Engineering Sciences Vol.2, Issue 1, Jan, 2012

[10] Yun Xue and Jianbin Chen, "Research and Design of Web Data Mining in Personalized E-Business" Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09), Nanchang, P. R. China, May 22-24, 2009, pp. 096-099

[11] Nivedita Roy and Tapas Mahapatra, "Web Mining: A Key Enabler in E-Business"@2005 IEEE.

[12] Sabyasachi Pattanaik,Partha Pratim Ghosh, "Role of Data Mining in E-Payment systems", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 2, February 2010