

WEB MINING – A CATALYST FOR E-BUSINESS

Abdul Rahaman Wahab Sait¹ and Dr.T.Meyappan²

¹Department of Computer Science, Shaqra University, Alquwaya,
Kingdom of Saudi Arabia
rahamaan@gmail.com

²Department of Computer Science and Engineering, Alagappa University, Karaikudi,
Tamilnadu, India
meyslotus@yahoo.com

ABSTRACT

In this world of information technology, everyone has the tendency to do business electronically. Today lot of businesses are happening on World Wide Web (WWW), it is very important for the website owner to provide a better platform to attract more customers for their site. Providing information in a better way is the solution to bring more customers or users. Customer is the end-user, who accessing the information in a way it yields some credit to the web site owners. In this paper we define web mining and present a method to utilize web mining in a better way to know the users and website behaviour which in turn enhance the web site information to attract more users. This paper also presents an overview of the various researches done on pattern extraction, web content mining and how it can be taken as a catalyst for E-business.

KEYWORDS

E-Business, Web Usage Mining, Web Content Mining, World Wide Web & Pattern Extraction

1. INTRODUCTION

Web mining is the subset of data mining, which works with the extraction of interesting knowledge from the WWW. Internet is a back bone for E-Business. It is a medium for the vendor to reach the customer and serve them in a better way to make them to revisit their site. Satisfying one customer gives more customers to the particular vendor. Data mining is the knowledge discovery technique from the huge amount of data. To mine the interesting user pattern from the huge pool of data, we can employ the web mining to improve the better understanding of the customer behaviour. This area of research is so huge today partly due to the interests of various research communities, the tremendous growth of information sources available on the web and recent interest in E-Business. Web mining field consists of main three categories, web usage mining, web structure mining and web content mining. In web usage mining the goal is to examine web page usage patterns in order to learn about a web system's users or the relationships between the documents. Web usage mining is useful for providing personalized web services, an area of web mining research that has lately become active. It promises to help tailor web services, such as web search engines to the preferences of each individual user [5].

E-Business is the electronic version of business which relies on internet. Today selling and buying are happening in web, even services also been done through internet. In fig 1 we gave the pictorial representation of E-Business. Enterprise resource planning online software is there to do outsourcing work. To attract the customers it is very important for the web site owner to provide the information in a different way from the other website. To do that work web mining can be taken as a tool to know the users and web behaviour.



Figure 1 – A Typical E-Business

As on Dec 2011, 135,676,044 [10] domains have been active in the web. It shows the important of the pattern extraction, which has taken from the web logs to improve the way of providing information to increase the number of customers. For example, if a company is having a website for service purpose or product selling purpose then they should know the customer behavior form the pattern or clustering.

2. RELATED WORK

S.K. Pani et al., presented an overview of web usage mining and also provided a survey of the pattern extraction algorithm used for web usage mining [1].

R.Cooley et al., presents an information about web usage mining and web content mining. They have studied different types of tools which are useful for pattern discovery. They proposed an overview on web content mining [2].

Chia – Hui Chang et al., proposed a pattern discovery approach to the rapid generation of information extractors that can extract structured data from semi – structured web documents. They have introduced a system for information extraction based on pattern discovery [3].

Rajni pamnani and Pramila chawan provided a survey and analysis of web usage mining systems and technologies. They also discussed about an application of an online recommender system that dynamically generates links to pages that have not yet been visited by a user [11].

Shohreh Ajoudanian and Mohammad Davarpanah Jazi proposed a method which new data mining algorithm that increases the speed of information matching. They have presented a mining algorithm that matches correlated attributes with smaller cost. Maintaining the Integrity of the Specifications [4].

3. WEB USAGE MINING

It is a process of extracting meaningful pattern from the web. In this section we are going to explain the mechanism of web usage mining. Data is very important, without data mining process will not begin. The sources of data are server access logs, referrer logs, agent logs and client side cookies. Client side cookies are not important source of data because those data stored in the client side [1][2][3], so it is fully depend upon the user or customer. They may delete the cookies after surfing the net.

Web server will take the entire request from the user and transfer it to the server. All the logging data will be found in web server log file. This file contains the entire request made to the web server, stored in a chronological order. A common log file format is the way to store the entire request made by the user.

Web usage mining studies reported to date have mined for association rules, temporal sequences, clusters and path expressions. As the approach in which the web is used continues to expand, there is a continual need to figure out new kinds of knowledge about user behavior that needs to be mined [6][11].

4. PATTERN EXTRACTION METHOD

Fig 2 shows the steps in pattern extraction. Those steps are the common method for the extraction of the pattern from the log files. They are

4.1. Data Collection

In section 3, we have explained the data collection and sources of data. Data collection on the WWW is incremental and also scattered in its very nature. Hence there is a need to develop mining algorithms that take as input the existing data, mined knowledge and the new data and develop a new model in an efficient manner.

If all the data were to be integrated before mining, a lot of valuable information could be extracted. However, an approach of collecting data from all possible server logs is non – scalable and impractical. Hence there needs to be an approach where knowledge mined from various logs can be integrated together into a more ample model [1][3].

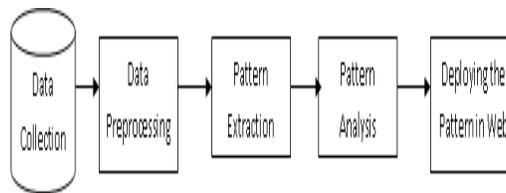


Figure 2 – Method of Pattern Extraction

4.2. Data Pre-processing

Web data are semi – structured. It contains mixture of much kind of data, so it is very noisy. It is necessary to give pre-treatment to carry on a unification transformation to those databases. Unwanted data like log image entries and robot assesses can be cleaned at this stage. Robot assess is nothing but a programmed access of information without using a human interaction.

4.3. Pattern Extraction

Extraction of pattern from the pre processed data can be done by using three categories. The three categories are Association rule mining, Clustering and Classification. Applying any one of the category we will get the pattern from the web log. Today we have lots of free web mining software and open source. Here it is not that much important to stress which category is best for the extraction.

4.4. Pattern Analysis

After the pattern extraction, there is a need to develop techniques and tools for enabling the analysis of discovered patterns. These techniques are expected to draw from a number of fields including statistics, graphics and visualization, usability analysis and database querying. Analyzing the pattern is very important to deploy the pattern in the site.

5. WEB CONTENT MINING

It is the process of information discovery from sources across the WWW. It is a form of text mining and can take advantage of the semi – structured nature of web page text [4][5]. To mine information from deep web that is a large set of dynamic queryable databases, we need a system that can extract automatically. For this purpose we use web content mining techniques and can take advantage of the semi – structured nature of web page text.

5.1. Agent Based Approach

The agent approach uses so called web agents to collect relevant information from the www. A web agent is a program that visits a web site and filters the information the user is interested in. There are three subtypes for the agent based approach: Intelligent search agents, information filtering / categorization and the personalized web agents.

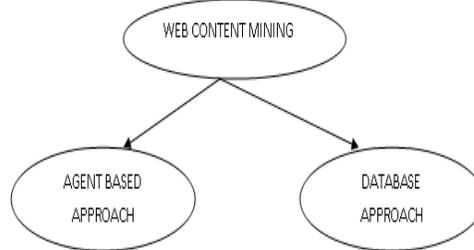


Figure 3 – Classification of Web Content Mining

5.2. Database Approach

The database approach for web mining tries to develop techniques for organizing semi – structured data stored in the web into more structured collections of information resources. Standard database querying mechanisms and data mining techniques can be used to analyze those collections. The database approach further classified into two types multilevel database and web query systems.

6. WEB MINING – AS A CATALYST

Web mining makes use of software agent to activate targeted offers as occurrence take place in real time. An agent is a collection of program that gains access on behalf of a process. It provides an enterprise the integrated tools to investigate all type of data resource from several departments from different position from different arrangements for a collection of deliverables such as tendency to procure scores and prediction of customer behavior. In the previous sections we have explained the methods involved in web usage mining and web content mining. It is necessary to know how these concepts can be employed to improve the business in the web. The fig. 4 shows the utilization of web mining as a catalyst for e-business.

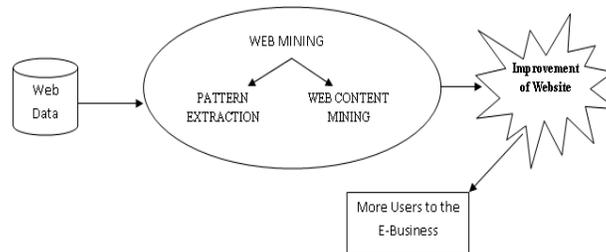


Figure 4 – Web mining – as a Catalyst

The above figure shows the pictorial view of the web mining process to help the e – business. Software agents exist to extract the information from the web data, using those software it is easy to know competitor web information. After drawing a clear picture of the competitor information, we can improve our website structure and presentation of information. This results in more number of users to the website. To promote e-business, we need a best recipe to provide information in an attractive way. Today a lot of businesses are happening on the web. Customers are very important for the vendor to survive in the business world. Many portals are there in the web for selling and buying the product. Everyone has to do something special to cover a group of people to bring more number of people to their business [7][8].

The two steps are in the process as follows:

1. Extracting the patterns from the competitors' site through web mining software agents.
2. Implementing the new changes in the website.

Using the web mining technique the organizations can enhance their own website in a better way. User behaviour is the important thing to know the stage of the business. By using pattern extraction method of web usage mining the website owner can know the user's behavior. To know the competitor position is also an important one to do a better enhancement of website, to do that web content mining technique exist. By using that the website owner can know the competitors site information.

Analyzing data from web logs can help organizations determine the life time value of customers, cross marketing strategies across products and effectiveness of promotional campaigns among other things. It can also provide information on how to restructure a website to create a more effective organizational presence and shed light on more effective management of workgroup communication and organizational infrastructure. For selling advertisements on the WWW, analyzing user access patterns helps in targeting advertisements to specific group of users. More sophisticated systems and techniques for discovery and analysis of patterns are now emerging in the market [1][9].

7. CONCLUSIONS

There is a budding trend among companies, organizations and individuals alike to gather information through web mining to utilize that information in their best interest. It is a demanding job for them to fulfill the user needs and keep their concentration in their website. This paper gave an idea about web mining and how it can be utilized in an efficient way to improve the business. Customer behavior is very important for an organization to enhance the way of providing information to attract them. Analysis of significant information will be helpful for organization to develop promotions that are more effective, internet accessibility, inter – company communication, structure and productive marketing skills through web usage mining. Pattern extraction and web content mining are the best tools to know the customer and web behavior.

REFERENCES

- [1] .K.Pani, L.Panigrahy, V.H.Sankar, Bikram Keshari Ratha, A.K.Mandal ,S.K.Padhi, “Web Usage Mining – A survey on pattern extraction from web logs”, Science, Vol 1, Iss. 1, 2011, PP. 15 – 23.
- [2] R.Cooley, B.Mobasher and J.Srivastava, “Web Mining: Information and Pattern Discovery on the World Wide Web”, Proc. IEEE International conference on Artificial Intelligence, 1997, DOI: 3 – 8 Nov 1997, PP.558 – 567.

- [3] Chia – Hui Chang, Chun – Nan Hsu and Shao – Cheng Lui, “ Automatic Information Extraction From Semi – Structured Web Pages by Pattern Discovery”, Decisin Support Systems 35 (2003), pp. 129 – 147.
- [4] Shohreh Ajoudanian and Mohammad Davarapanah Jazi, “ Deep Content Mining”, Science, 2009, PP: 501 – 505.
- [5] Miguel Gomes Da Costa Junior and Zhiguo Gong, ”Web Structure Mining: An Introduction”, Proc. IEEE conference on Information Acquisition, June 27 – July 3, 2005, pp. 590 – 595.
- [6] Chen Hu, Xuli Zong, Chung – Wei Lee and Jyh – haw yeh, “ World Wide Web Usage Mining Systems and Technologies”, Science,2003 Vol. 1, No. 4, pp. 53 – 59.
- [7] S.Rawat, L. Rajamani, “Discovery Potential User Browsing Behaviors Using Custom – Built Apriori Algorithm”, Science (IJCSIT) Vol.2, No.4, August 2010, pp.28 – 37, DOI. 10.5121 / ijcsit 2010.2403.
- [8] B.Chidlovskii, V.Borghoff, P.Chevalier, ”Towards Sophisticated Wrapping of Web Based Information Repositories”, Proc. Of the 5th International RIAO Conference,1997, pp. 123 -125.
- [9] R.B.Doorenbos, O.Etzioni,D.S.Weld, ” A Scalable Comparison – Shopping Agent for the World Wide Web”, Proc. Of the 1st International conference on Autonomous Agents, 1997, pp.39 – 48.
- [10] [http:// www.domaintools.com/internet - statistics/](http://www.domaintools.com/internet-statistics/)
- [11] Rajni Pamnani and Pramila Chawan, “ Web Usage Mining: A Research Area in Web Mining”.

Authors

1. ABDUL RAHAMAN WAHAB SAIT was born in Apr. 19, 1981 and he has completed his Masters in Information technology in 2003 in Madras University, India. Later he has done his Master of philosophy in computer science in 2007 at Periyar university, India. Now he is working as a Lecturer in Computer science, Alquwaya, Shaqra University, Kingdom of Saudi Arabia. He got interest in Virtual reality and Artificial Neural Networks. He has written many computer articles and presented paper in National conference in India



2. Dr. T.Meyyappan M.Sc,M.Phil,M.B.A currently, Associate Professor, Department of ComputerScience and Engineering, Alagappa University, Karaikudi, TamilNadu. He has published a number of research papers in National and International journals and conferences. He has developed Software packages for Examination, Admission Processing and official Website of Alagappa University. As a Co-Investigator, he has completed 1.70 crore project on smart and secure environment funded by NTRO, New Delhi. He has been honoured with Best Citizens of India Award 2012. His research areas include Operational Research, Digital Image Processing, Fault Tolerant computing, Network security and Data Mining.

