International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.6, No.6, June 2017 DOI:10.15693/ijaist/2017.v6i11.433-436

Role of Web Mining: Classification and Tools in E-Commerce

J.Priya[1], M. Sreedhar reddy[2]

[1]Assistant Professor, [2]Professor
Department of Computer Science and Engineering,
Samskruti College of Engineering and Technology,
Hyderabad.

Abstract: The source of information for E-Commerce is huge volume of structured and unstructured data which is called big data. Actually those data are collected from customer's internal processes, markets and business environment. Web mining is the mixture of information collected by usual data mining tools and techniques with information gathered from the Web. The Web is collection of interconnected files on one or more Web Servers. Actually Web mining is a part of data mining, which helps to find out new patterns that make possible to take new decisions by a merchant for business improvement. In this paper we discuss about applications, processes of web mining techniques, web mining tools and the major challenges in the perception of E-commerce.

Keywords: Web mining, Data mining, E-Commerce, Web data

1. Introduction

Data mining, *the extraction of concealed information by using various patterns from vast databases*, is a dominant new technology with great potential to help organizations spotlight on the most important information in their data warehouse. Data mining tools is used to predict future trends and behaviors, permits organizations to make optimistic and knowledge-driven decisions.E-commerce is doing commercial business through the Web. That means selling and buying products through online payment or cash on delivery over internet is called E-Commerce. Actually these two operations take place among either customer – seller or seller – seller. Nowadays E-Commerce is preferred by most of the buyers because of its easy accessibility, easy return policy, trust on seller and universal delivery.Web Mining is one of the applications of Data Mining techniques, which is used to extract remarkable and useful patterns and concealed information from web documents and web activities [1]. This information contains customer's activities such as frequently bought items and interested items. Through this information the seller can send advertisement or e-mail about the related products to the appropriate customers. By using web mining the sellers can improve their understanding about their consumers and easily they can improve their selling percentage too.

2. Classification of Web Mining

Web mining is the mixture of information composed by usual data mining technique and methodologies with information gathered through the internet. That information is collected for provide better customer service and improve sales. Web mining is classified into major three types: web content mining, web structure mining, and web usage mining. It is shown in the following figure.

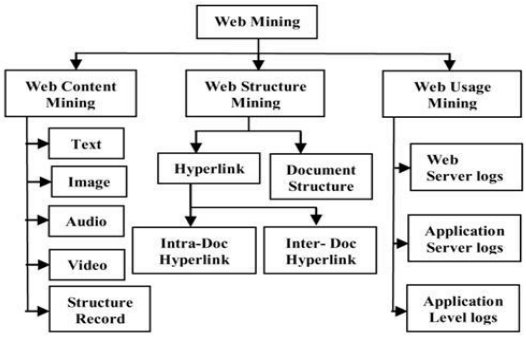

Figure 1: Classification of web mining

Web content mining is used to mining the essential information from web pages. This information may be any of the following text, image, audio, structure record and video. Some one can easily misunderstood that both data mining and web content mining are identical. But data mining is deal with structured information and web content mining is deal with unstructured and/or semi-structured data. Web content mining involves in three techniques named summarization, classification and clustering of web contents. It provides useful pattern about users needs and customer behavior. The tools used for web content mining are screen scaper, mozenda, automation anywhere, web content extractor, web info extractor. For example, while customer try to search some product through some keyword then system will automatically shows some choices in search engine.Web structure mining or link mining is used to identify the useful information from the structure or graph which is generated by using the links among more than one web page. This graph is

also called web graph which contains web pages as nodes and hyperlinks as edges. This web structure mining is further divided into two types depends on type of structure they are extracting patterns from hyperlinks and mining the document structure. Web content is mainly focusing on content of the web pages but web structure mining is focusing on the links between the web pages. The efficient techniques for implementing web structure mining are page rank algorithm and HITS algorithm.Web usage mining or web log mining is used to discover the user navigation patterns from the web log that means predict the user search. That means identifies browsing patterns by user navigational behavior. The tools used for web usage mining are categorized under two tools named as pattern analysis tool and pattern discovery tool. The goals of link mining are improves system performance, improves the design of an e-commerce web site, enhance the security of system and provides support for marketing design. The most useful data mining techniques for web usage mining under these two tools are sequential pattern analysis, clustering, classification, path analysis, OLAP, grouping and association rule mining.

3. Phases of web usage mining

Web usage mining is a complete process that includes various steps of data mining cycles that includes web data pre-processing, pattern discovery and pattern analysis. After these three steps the user behavior or user access pattern can be able to identify. This procedure is shown in the following figure.

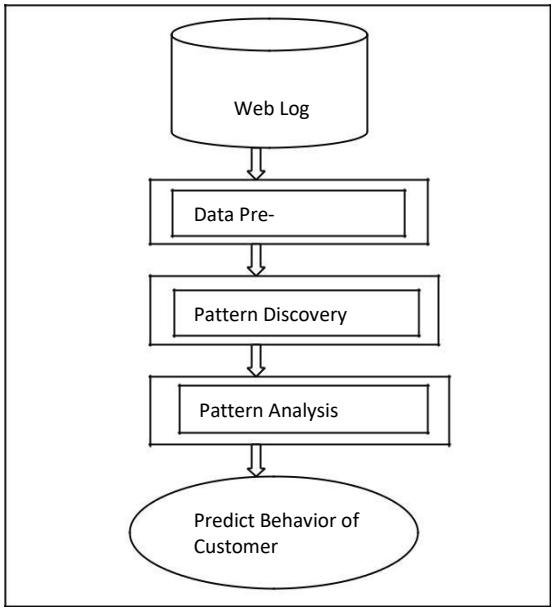

Figure 2: Phases of web usage mining

a) Web Log or Web Data

Web usage mining is used to predict behavior of customers. To predict that, the information such as web pages accessed, time of access, spent time, next link web page etc. by the particular user collected from servers using traditional data mining techniques [2]. These data are categorized under the following four categories [3].

- Content data which contains images, text or any simple data retrieved from database.
- Structure data which represents the way of content is organized. Those may be either HTM/XML tags or hyperlinks.
- Usage data contains user's access details such as access time and date, IP address, path to this page etc.
- User profile data which provides the information about the user such as preferences, interests etc.

b) Web Data Preprocessing

Once the data have been collected the next step is to pre-process. Actually this pre-processing step is essential to provide the accurate web data [5]. Pre-processing is performed through five steps named data pre-processing, cleaning of data, identification of user, identification of user session and identification of transaction.

c) Pattern Discovery from Web Data

By applying classic data mining methods such as path analysis, clustering analysis, association rule, sequential patterns, dependency modeling and classification rule can able to detect some interesting patterns. Those patterns give us useful knowledge about the web data [2].

d) Analysis of Pattern of Web Data

Once the patterns have been discovered those should be evaluated for eliminating uninterested or unwanted pattern by using tool called OLAP. By discovers interesting pattern the organization can able to provide a better customer service.

4. Use of Web mining in E-commerce

Among all of the possible applications of web mining e-commerce takes an important place [8]. So any details regards consumers got a highest preference in e-commerce buyer's market. Web mining techniques and types of web mining are used to understand the consumer's requirements [4]. By the better understanding of consumer the merchant can hold a strong place in competitive e-commerce era [7]. At this

point we can say that e-commerce is preferred, by more number of the consumers compared to past ten years, because of its low cost, convenient, easy access and safe. The area of applications of web mining in E-Commerce is summarized in the following diagram.

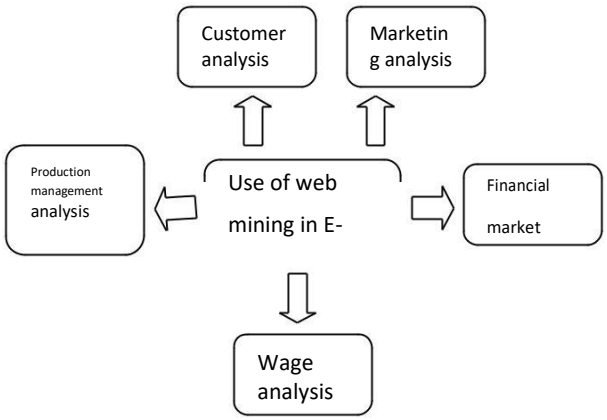

Figure 3: Use of web mining in E-Commerce.

Financial market Analysis:
The Financial market analysis is one of the difficult processes. Because by using irregular data and noisy data the analyst should predict the future market. Even though it is difficult this can be achieved by using the comparative analysis of organizations profit statement, analysis of income statement, analysis of corporate balance sheet and profitability and controlling over them. Web mining tool can achieve all these analysis easily [9].

Customer Analysis:
Customer analysis is one of the important processes to improve the organization's profit. So this analysis has the highest priority among all criteria. Using web mining tools the organization can improve the customer satisfaction by better understanding of every customer [9].

Marketing Analysis:
It contains profit of sales, profit margins, targets of sales, time to take delivery and improvement of products against competitors. By using web mining tool these all can effectively analyze.

Wage analysis:
Wage analysis which includes employment types, paying methods, surcharges for employees, includes detection, incentives etc., and analyze of average wage [9].

Production management analysis:
This includes stock of items, details of sold items, product return policy and product's dynamicity etc., Which automatically increase product selling rate [6].

5. Web Mining Tools in E-commerce

Web mining tool is one of the software components which apply some data mining techniques to discover or determine some useful patterns from large data sets to improve the merchant's sales percentage in a market. To get better understanding of customers, sales market etc. any one of the web mining tools has to be applying on the web data. The overall process of technique and tool is shown in the following diagram.

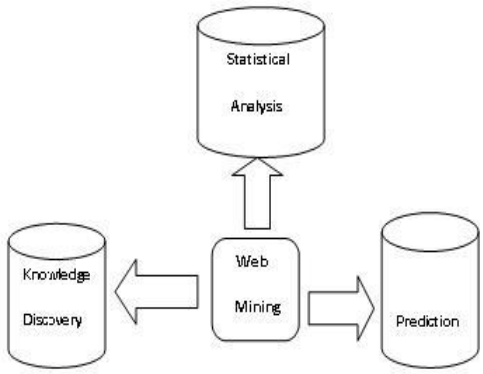

Figure 4: Process of Web Mining

A) Octoparse:

Octoparse is a simple but very powerful tool for web mining to extract web data from the World Wide Web. Mainly this type of tool is used for Web Content Mining and it can installed in Windows operating system. By using this, merchants can extract accurate information.

B) Oracle Data Mining (ODM):

ODM is software which is developed by Oracle. Oracle Data Mining processes use built-in-
Features of Oracle Database to maximize scalability and make efficient use of system resources.
This tool can use in Windows and it is the tool for Web Usage Mining.

C) Tableau:

Tableau is interactive tool which is mainly focused on business intelligence. It finishes the process only in seconds or minutes and is achieved through the use of an easy to use drag-and-drop interface. It is the tool for Web Usage Mining. It can be used either in Mac or Windows.

D) Scrapy:

Scrapy is an open source tool for gathering data from websites. It is developed in Python and we can write the rules to extract web data. It is supported in either Linux

or Windows or Mac or BSD.
This is ultimately developed for Web Content Mining.

E) HITS Algorithm:
Hyperlink Induced Topic Search is abbreviated as HITS. It retrieves the pages based on the links between the pages. The search is based on the search query. This tool is developed for Web Structure Mining.

F) Page Rank Algorithm:
It is a famous web structure mining algorithm. It is used to calculate the weightage of the web elements.

6. Challenges in Web Mining

- As information in web is extremely enormous and rapidly growing it turning to a very big challenging task to mine the data from the web.
- It turns out to be hard to handle unstructured, unusual, heterogeneous and asymmetrical data patterns.
- Systematize hardware and software for such a hard and tremendously large processing is also not easy.
- Even the source of the information is changing rapidly time to time which gives challenge for web mining.
- Web mining tools have to ensure that the requirements of users from different backgrounds who are having a different request with unique interest are satisfied.
- The user can efficiently retrieve required information from web but providing retrieval of information with privacy is a greatest challenge for web mining.
- One of the real time challenges for web mining is fraud and threat analysis.

7. Concluding remarks

Web Mining is a rapidly developing field in E-Commerce. Nowadays in E-Commerce the vast amount of information has to use. To organize those information web mining plays an important role. The big retailer uses these data for predicting the user requirements and interest. So that, they can attract the consumers by sending interesting searches to them while shopping on the web site. This paper provides the detail study of types of web mining, role of web mining and tools of web mining in the perspective of E-Commerce.

References

1) Li Mei , Feng Cheng , "Overview of WEB Mining Technol ogy and Its Application in E-commerce", 2010, IEEE.
2) International Journal of advancements in Research & Technology, Volume 1, June-2012
3) ISSN 2278-7763 "Efficient Preprocessing technique using Web log mining" by Sheetal A. Raiyani et.al.
4) Chen L, Sycara K (1998) A Personal Agent for Browsing and Searching. In Proceedings of the 2nd International Conference on Autonomous Agents, Minneapolis/St. Paul, May 9-13, pp132-139
5) BVIMSR'S Journal of Management Research ISSN: 0976-4739 "Web Mining techniques, process and applications in Ecommerce" Prof. Rajesh R. Gawali et al. April 2013 , pp-19 -20.
6) International Journal of Electronics and Computer Science Engineering ISSN- 2277-1956 "Web Mining: Characteristics and Application in E-Commerce" by Jyoti Yadav et al.
7) International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2015 "Role of Web Mining in E-Commerce" by Arti et al. pp-251-253.
8) Int. J. Communications, Network and System Sciences, 2015, "Data Mining in Electronic Commerce:Benefits and Challenges" by Mustapha Ismail et al. pp- 501-509
9) "Web Mining Applications in E-Commerce and E-Services" by I.-H. Ting, H.-J. Wu (Eds.)Studies in Computational Intelligence, Vol. 172 , pp 10-20.
10) Jyoti Yadav, Bhawna Mallick," Web Mining: Characteristics and Application in Ecommerce", International Journal of Electronics and Computer Science Engineering, ISSN- 2277-1956, ISSN 2277-1956/V1N4-2020-2025.