

Web Mining Role in E-Commerce : Classification and Tools

Nimisha Sajad

Assistant Professor, Department of Computer Science, University Institute of Technology, Trivandrum, India

To Cite this Article

Nimisha Sajad, "Web Mining Role in E-Commerce : Classification and Tools", *International Journal for Modern Trends in Science and Technology*, Vol. 06, Issue 01, January 2020, pp.-53-57.

Article Info

Received on 03-December-2019, Revised on 26-December-2019, Accepted on 06-January-2020, Published on 23-January-2020.

ABSTRACT

The source of information for E-Commerce is vast amount of organized and unstructured data which is called big data. Actually those data are obtained from customer's internal processes, marketplaces and business environment. Web mining is the mixing of information obtained by standard data mining tools and methodologies with information received from the Web. The Web is collection of linked files on one or more Web Servers. Actually Web mining is a part of data mining, which helps to search out new patterns that enable feasible to take new decisions by a merchant for commercial progress. In this article we speak about applications, processes of web mining techniques, web mining tools and the significant obstacles in the perception of E-commerce.

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

Copyright © 2015-2020 International Journal for Modern Trends in Science and Technology
All rights reserved.

I. INTRODUCTION

Data mining, which is the extraction of hidden information from enormous databases via the application of diverse patterns, is a dominating new technology that has the ability to assist companies in focusing on the most critical information contained inside their data warehouses. It is possible to foresee future trends and behaviours through the use of data mining techniques, which allows companies to make optimistic and knowledge-driven decisions.

E-commerce is the practise of conducting commercial transactions through the Internet. E-Commerce is the term used to describe the act of selling and purchasing goods over the internet using online payment or cash on delivery. In reality, these two transactions take place between either a consumer and a vendor or a seller and a customer. The ease of accessibility, the simple return policy, the faith placed in the seller, and the

availability of worldwide shipping are all reasons why E-Commerce is becoming increasingly popular among purchasers.

[1] Web Mining is an application of Data Mining methods that is used to extract amazing and valuable patterns and hidden information from web publications and web activity. This information provides information on the behaviours of customers, such as regularly purchased things and those that have piqued their interest. The merchant can use this information to send advertisements or e-mails about relevant items to the consumers who have expressed interest in them. Sellers may have a better knowledge of their customers with the use of web mining, and they can easily increase their sales % as well.

II. CLASSIFICATION OF WEB MINING

Web mining is the blend of information created using standard data mining methodology and techniques with information collected from the

internet. That information is gathered in order to provide better customer service and increase sales opportunities. Web mining may be divided into three types: web content mining, web structure mining, and web use mining. Web content mining is the most common sort of web mining. It is depicted in the following illustration.

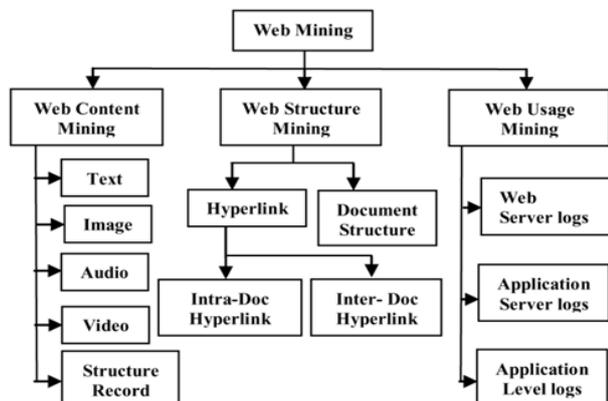


Figure 1: Classification of web mining

Web content mining is a technique for extracting the most important information from web pages. Text, image, audio, structure record, and video are all acceptable formats for this information. Those unfamiliar with the field may believe that data mining and web content mining are interchangeable terms. On the other hand, data mining deals with organised information, whereas online content mining deals with unstructured and semi-structured material. Three strategies are used in online content mining, including summary, classification, and clustering of web-based material. It gives a valuable pattern regarding the demands of users and the behaviour of customers. Web content mining is accomplished through the use of tools such as screen scaper, mozenda, automation anywhere, web content extractor, and web info extractor. For example, if a customer attempts to search for a product using a specific keyword, the system will automatically display a number of options in the 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, which is also known as a web graph, is made up of nodes that are web pages and edges that are hyperlinks. This web structure mining is further subdivided into two types based on the type of structure they are mining: extracting patterns from hyperlinks and mining the structure of the document itself. Online content mining is primarily concerned with

the content of web sites, whereas web structure mining is concerned with the links that connect the web pages. The efficient techniques for implementing web structure mining are page rank algorithm and HITS algorithm.

Web usage mining, also known as web log mining, is a technique for discovering user navigation patterns from web logs, which can then be used to predict the user search. That is, it identifies browsing patterns based on the navigational behaviour of the user. The tools that are used for web usage mining are divided into two categories: pattern analysis tools and pattern discovery tools. Pattern analysis tools are used to identify patterns in web usage data. 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 sequential pattern analysis, clustering, classification, route analysis, OLAP, grouping, and association rule mining techniques are the most helpful data mining approaches for online usage mining under the umbrella of these two instruments.

III. PHASES OF WEB USAGE MINING

It is a comprehensive process that includes several steps of data mining cycles, such as web data pre-processing, pattern discovery, and pattern analysis, among others. After these three processes the user behaviour or user access pattern can be able to identify. An illustration of this process may be seen in the accompanying figure.

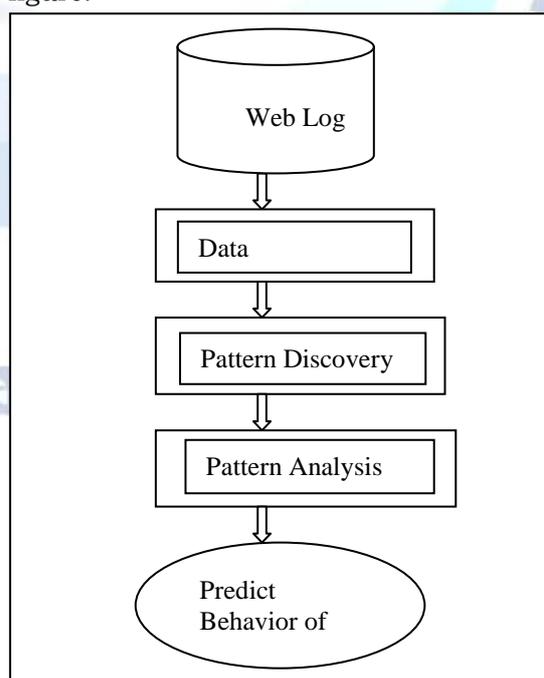


Figure 2: Phases of web usage mining

Web Log or Web Data

Web use mining is a technique for predicting customers' behaviour. Traditional data mining techniques [2] were used to anticipate this based on information acquired from servers, such as web sites viewed, time of access, time spent, next link web page, and so on, by a specific user. These data have been classified into four groups [3], which are as follows:

Web Data Preprocessing

Data pre-processing is carried out following collection of the required data sets. It is actually necessary to do this pre-processing step in order to offer reliable web data [5]. A total of five stages are required to complete pre-processing, including data preparation, data cleaning, identification of the user, identification of the user's session, and identification of the transaction.

Pattern Discovery from Web Data

Through the use of conventional data mining approaches such as route analysis and clustering analysis (as well as the association rule), sequential patterns (as well as dependency modelling and classification rule), it is possible to discover some intriguing patterns. Those patterns provide us with valuable information about the web data [2].

Analysis of Pattern of Web Data

Once the patterns have been found, they should be analysed in order to eliminate any uninteresting or undesired patterns, which may be done with the use of an OLAP tool. The firm may be able to give superior customer service as a result of the discovery of an intriguing trend.

IV. USE OF WEB MINING IN E-COMMERCE

E-commerce is a significant application of web mining, and it has a prominent position among its conceivable uses [8]. As a result, any information pertaining to customers received the highest level of priority in the e-commerce buyer's market. Web mining methods and forms of web mining are employed in order to better understand the requirements of the customer [4]. A deeper knowledge of the consumer can help the merchant maintain a solid position in the competitive e-commerce environment [7]. At this moment, we can state that e-commerce is favoured by a greater number of customers now than it was 10 years ago, owing to its low cost, convenience, ease of access, and safety, among other factors.

The following figure depicts a high-level overview of the uses of web mining in the field of e-commerce.

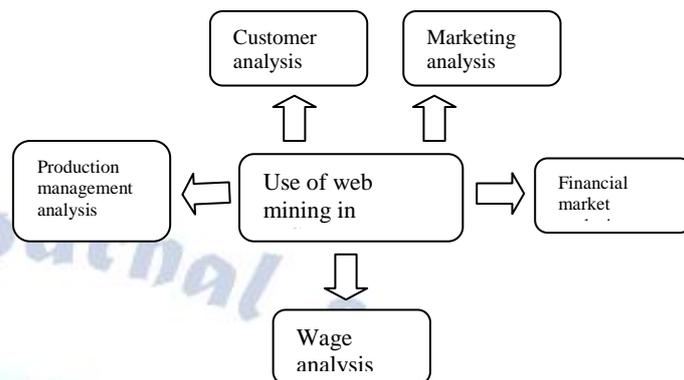


Figure 3: Use of web mining in E-Commerce

Financial market Analysis:

The examination of the financial markets is one of the most challenging tasks. Because the analyst should be able to estimate the future market by utilising irregular data and noisy data. Even if it is challenging, this may be accomplished through the use of comparative analysis of companies' profit statements, income statements, corporate balance sheets, and profitability, as well as the exercise of control over them. All of these analyses may be accomplished with relative ease using a web mining programme [9].

Customer Analysis:

Customer analysis is one of the most significant activities for increasing the profitability of a firm. As a result, this analysis takes precedence over all other considerations. [9] By utilising web mining techniques, a firm may increase customer satisfaction by gaining a deeper understanding of each individual consumer.

Marketing Analysis:

A profit of sales report comprises information on the profit margins, sales objectives, delivery timeframes, and product improvement in comparison to the competition. These may all be successfully analysed with the help of a web mining programme.

Wage analysis:

Employment kinds, payment systems, surcharges for employees, detection, incentives, and other factors all contribute to wage analysis [9], which also includes an examination of the average salary.

Production management analysis:

This comprises supply of products, specifics of sold items, product return policy, and the dynamic nature of the product, among other things, all of which help to raise the product selling rate [6].

V. WEB MINING TOOLS IN E-COMMERCE

It's one of the software components that uses data mining techniques to discover or determine some valuable patterns in vast amounts of information to help a merchant increase his or her share of the market. Web mining tools are used to help merchants increase their share of the market by increasing their sales percentage. Any one of the web mining technologies must be used on online data in order to gain a better knowledge of clients, sales markets, and other aspects of business. The following diagram depicts the whole process of using a method and an instrument.

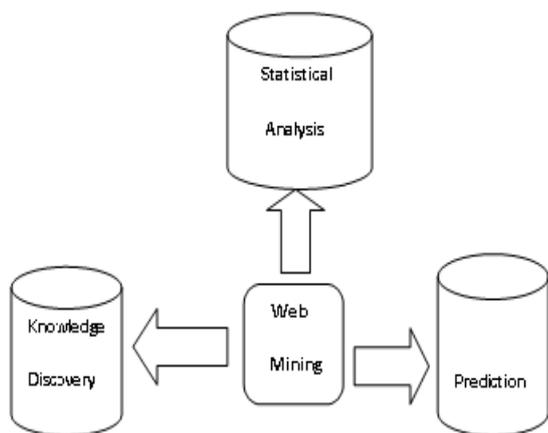


Figure 4: Process of Web Mining

Octoparse:

Octoparse is a straightforward yet extremely effective web mining application that allows you to extract web data from the World Wide Web. This sort of programme is mostly used for Web Content Mining, and it may be installed on computers running the Windows operating system. Merchants may obtain correct information by utilising this method.

Oracle Data Mining (ODM):

Oracle Development Manager (ODM) is a piece of software created by Oracle. Data mining procedures in Oracle Database make advantage of built-in capabilities to increase scalability and make more effective use of system resources. This programme is available for Windows users, and it is used for Web Usage Analysis.

Tableau:

Tableau is an interactive business intelligence solution that is primarily focused on business intelligence. The procedure is completed in seconds or minutes thanks to the usage of a simple drag-and-drop interface that is straightforward to understand and navigate through. It is the Web Usage Mining tool that is used. It may be used on either a Mac or a Windows computer.

Scrapy:

Scrapy is a free and open source programme for collecting information from the internet. It is written in Python, and we can design the rules that will extract web data from the system. It may be used with either Linux or Windows, as well as Mac OS X or BSD.

This is eventually intended for use in Web Content Mining applications.

HITS Algorithm:

HITS stands for Hyperlink Induced Topic Search, which is an abbreviation for the term. It fetches the pages based on the linkages that exist between the pages in the database. The search is conducted in accordance with the search query. This tool is being developed for the purpose of Web Structure Mining.

Page Rank Algorithm:

It is a well-known method for web structure mining. It is utilised in the calculation of the weighted average of web components.

VI. CHALLENGES

- 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.

VII. CONCLUSIONS

A fast emerging discipline in the realm of e-commerce is web mining. In today's world of E-Commerce, a great amount of information must be utilised. A significant role is played by web mining in the organisation of such information. These data are used by the large store to forecast consumer requirements and interests. In order to attract customers by giving them relevant queries while they are shopping on the website, they must first grab their attention. This article presents a detailed examination of the many forms of online mining, the function of web mining, and the web mining technologies from the standpoint of e-commerce.

REFERENCES

- [1] Li Mei , Feng Cheng , "Overview of WEB Mining Technology and Its Application in E-commerce", 2010, IEEE.
- [2] International Journal of advancements in Research & Technology, Volume 1, November-2012 ISSN 2278-7763 "Efficient Preprocessing technique using Web log mining" by Sheetal A. Raiyani et.al.
- [3] 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
- [4] BVMSR'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.
- [5] International Journal of Electronics and Computer Science Engineering ISSN- 2277-1956 "Web Mining: Characteristics and Application in E-Commerce" by Jyoti Yadav et al.
- [6] 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.
- [7] Int. J. Communications, Network and System Sciences, 2015, "Data Mining in Electronic Commerce:Benefits and Challenges" by Mustapha Ismail et al. pp- 501-509
- [8] "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.
- [9] 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.