

WEB DATA MINING IN E-SERVICES – CONCEPTS AND APPLICATIONS

Sunil

Department of University Polytechnic
Faculty of Engineering & Technology
Jamia Millia Islamia, New Delhi, India
sunil@jmi.ac.in

Prof. M. N. Doja

Department of Computer Engineering
Faculty of Engineering & Technology
Jamia Millia Islamia, New Delhi, India
mdoja@jmi.ac.in

Abstract

With the rapid development of internet and its efficiency and convenience people prefer online services that results a huge amount of data increasing day by day. In a short span of time e-services have observed a tremendous growth worldwide. A large amount of data which is being produced by the e-services during transaction is much unpredictable but contains hidden and very useful information which could help organizations to find more potential customers', increase trading volume, improve product publicity and restructure their websites. In this regard this paper discusses the web data mining strategies and applications in e-services which are required for optimizing website structure that will help for business and learning platforms to increase their revenues, attract new customers or learners, retain the old one and help developers to increase the frequency of learners' visits.

Keywords: E-services, E-learning, E-commerce, Web data mining.

1. Introduction

World Wide Web now a day is broad extend medium for the flow of information. As the rapid expansion in the technology the quantity of data and its versatile structure are growing rapidly. There comes an urge for web data mining. A quick development in online services called e-services[4], its functions like e-learning, e-government, e-commerce, e-market, e-finance, e-banking etc. has made trade society and consumers face new circumstances.

E-learning being a structure consisting of all form of electronically supported learning and teaching at anytime and anywhere. Communication and information plays an important role in implementing the learning process [10]. E-learning is a critical support mechanism for educational institutions so as to grow the performance of their students, teachers, as well as useful for organizations to enhance their performance [10, 1].

The E-commerce sector is booming in various dimensions [6]. It is considered as bright and fastest growing sector worldwide. Business transactions over internet are considered as popular, comfortable and reliable now a days because of their well structured and customer friendly websites. Hence the real challenge for the industry is to focus on the effective and optimized website. Therefore it is very important to devise new marketing strategies, one of the effective solutions is recommendation system which provides and guides the customers that which type of product they want and learners about what kind of information in which they are interest [7].

2. Overview of E-services: E-learning and E-commerce

E-services are the online services like E-learning, E-government, E-commerce, E-market, E-finance, E-banking etc. This paper spotlights on function of web mining in two major services i.e. E-learning and E-commerce.

2.1. E-learning

E-learning is becoming a vital tool to support the learning system to accomplish its goals. E-learning became searing area in the 1990's after the spread of the Internet. Although it has a relative short history, it is becoming an imperative part of the learning [10, 3]. The majority of the universities adopted some kinds of E-learning within their learning system. E-learning may either be synchronous or asynchronous [5].

2.1.1. Synchronous E-learning

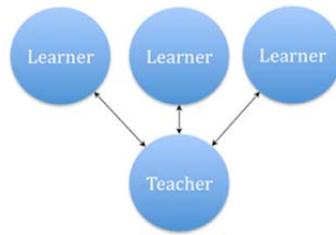


Fig. 1. Synchronous E-learning

In synchronous E-learning the information is accessed instantly as the communication occurs at the same time between individuals. It is characterized by instantaneous, real time and two way communications. This is commonly supported by media such as virtual classrooms, audio/video conferencing and chat sessions.

2.1.2. Asynchronous E-learning

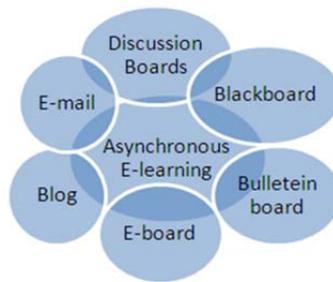


Fig. 2. Asynchronous E-learning

Asynchronous E-learning is usually facilitated by media such as e-mail, discussion boards, supports work relations among learners and with teachers, even when participants cannot be online at the same time. It is a key component of flexible learning. Thus it means that it need not be online at the same time in asynchronous E-learning. Thus E-learning makes possible for learners to log on to an E-learning environment at any time and download documents or send messages to others.

2.2. E-commerce

Buying and selling of commodities or services through electronic media such as internet and other computer networks is termed as E-commerce [2]. At first the term was applied to the execution of transactions through electronic transactions such as electronic data exchange. However in the mid of 90's, with the arrival of internet it began mainly referring to the sale of commodities and services on the internet, mainly using electronic payment. Since the hasty spread of internet the amount of trade conducted electronically has grown extremely. A high variety of business is made in this way as motivating the creation and use of innovations such as electronic fund transfer, marketing on internet, electronic data exchange, inventory management and automated management of data.

Table 1. Comparison of web data mining process questions in e-learning and e-commerce.

E-learning	E-commerce
Who are the students with highest frequency of logging-in?	Who are my most profitable customers?
Who are students most engaged on discussion boards?	Who are my repeat website visitors?
What pages do the students access most?	Who are my loyal customers?
What kinds of students are likely to get a high score online?	What clients are likely to defect to my rivals?

3. Process of Web Data Mining

The processes of web data mining includes-- data collection, data pre-processing, data storage, data mining and the final result we get is the knowledge discovery[13].

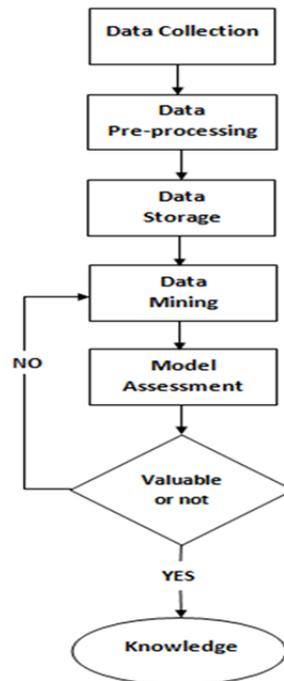


Fig. 3. Process of web data mining

3.1. Data collection

The first step is to collect the data on which we want to discover the hidden information. The data source of the e-commerce website includes customer personal information; customer personal information includes not only the customers' registered personal data but also includes its order information, customers' individual requirement and problems. Also data is collected about customers browse records and visited pages and behaviour information that help to analyse customer preference so as to forecast the future purchase behaviour of the customers. The second source is the server information that is generated on the server when client access the server, the data includes server log, error logs, agent logs files, cookie logs and transaction database.

3.2. Data pre-processing

As after the data is collected as log files the pre-processing is being done. The data pre-processing includes data cleansing, user identification and transaction identification.

3.2.1. Data cleansing

In data cleansing the unwanted data or we call errors are removed and the deletion of redundancy in data is done. The data cleansing done on multimedia files, java applet files, java script files, CSS files, pop-up ads and error access records.

3.2.2. User identification

In this every user in being identifies from the use log files. The most common technique used is based on the idea of heuristic rules as the different IP addresses represents different users, although id and the IP addresses are same then the agent log files can figure out whether the user is same or not by detecting the browser and operating system, if there is a condition that IP address, browser and operating system are same, then the user can be identified by page link structure.

3.2.3. Transaction identification

In this the users' behaviour is defined from their records. Here first each record is described then identified sessions of every user visit is organised in chronological order.

3.3. Data Storage

After data pre-processing is done the data now stored in the database in order to get ready to be extracted and used. Here in web data mining transactional database is used as compared to relational database.

3.4. Data Mining

After pre-processing and data storage we choose different data mining technologies that are based on the different requirements of the users. There are many data mining technologies which are association rules, clustering, classification and sequential patterns.

3.4.1. Association rules

By applying the association rule generation, in the web domain the pages which are largely often referenced together can be put in one single server session. To discover unordered relationship between items found in a database of transactions, association rule mining technique can be used[9]. In terms of web usage mining the association rules refer to a set of pages that are being accessed together with a support value that exceeding some specified threshold[8]. With the presence or absence of the association rules the web designers can restructure their web sites efficiently.

3.4.2. Clustering

It is a technique to group together the data items or users with similar characteristics. Clustering of user information or data items can facilitate the expansion and carrying out of future marketing strategies [9]. The users who have similar navigation patterns, group of those users can be revealed by clustering.

3.4.3. Classification

It is the technique of mapping a data item into one of the several predefined classes. In the web domain if users belonging to a particular classes or category, by using this technique web master can create a profile of such users. This requires extraction and selection of features that best describes the properties of a given class or category. By using supervised inductive learning algorithms such as decision tree classification, k-nearest neighbour classifier, support vector machine, naïve Bayesian classification etc, the classification can be done [8].

3.4.4. Sequential Pattern

The sequential pattern techniques aim to find the inter-session pattern, as a set of the items follows the presence of another in a time-ordered set of sessions. For the web marketer it is very meaningful to predict the future trends.

3.5. Model Assessment

When the system has implemented the algorithm, we get a huge number as thousands of models or rules. Only a small part of them is valuable and interesting. So for that the explanation and assessment is essential. To understand those complex data through using computer as to create some visible images, the technology used is visualization. As some decision makers are not so expert in data mining, that's why these models are required to convert into the forms as they are easier to understand and used by the users to make decisions.

4. Application of Web Data Mining in E-services (E-learning and E-commerce)

As E-services are the online services like E-learning, E-government, E-commerce, E-market, E-finance, E-banking etc. Here we are spotlights the application of web data mining in two major services i.e. E-learning and E-commerce.

4.1. Application of Web Data Mining in E-learning

Web data mining is the incorporation of data mining and web technology. In E-learning by mining web server log records, the web data mining can discover user access patterns, such as which pages are most visited by the learners, the number of learners visited the page, the places from where they visited and what is the possible

next page the user will visit after visiting the current page. The applications of web data mining in E-learning are as follows [12, 14]: Personalized service, system improvement, site structural modifications, intelligent service and personalized recommendation.

4.1.1. Personalized service

It will accumulate and categorize the user's access patterns. In accordance with the users current model to access the site it will be categorize. The personalized services will be provided by attaching pages dynamically to the visited current page of the user with the other pages chosen by the other users. By mining the single learners' browsing information, its visiting frequency, contents length of visit, it is used to find the learners' interest and preferences.

4.1.2. Personalized service

For measuring the customer contentment, the performance of web services and service quality are the key indicators. For an improved design of the web server, load balancing, data distribution, page caching and network transmission, web visiting information mining can provide valuable knowledge.

4.1.3. Site structural modifications

For attracting learners to learn, the structure and contents of the site plays a major role. With the help of web usage mining the designers of the site can improve the structure of the site by mining the visitors' feedback and its navigational path, as which page is mostly accessed by the user and how the pages are linked together. A well designed website is always considered as a successful website [11]. It must be designed such a way that it can always satisfy the needs of the users. An optimized, reorganized and structured website always attracts learners. In an E-learning context a recommender system is a software agent that tries to smartly recommend actions to a learner based on the actions of the previous learners.

4.1.4. Intelligent service

The services provided by the asynchronous e-learning platforms it should be added more intelligent factors so as to understand the learners learning intent by mining their visiting behaviour. The intelligent information service can identify and predict the users' interest and hobbies according to their browsing pattern and information inquiries.

4.1.5. Personalized recommendation

When a learner or user visits a website for their information, they are interested in some pages of their interest but they have to visit all pages to find the desired information. During their visit the server logs a record of their visit and after mining through that logs, interest of the user can be evaluated. The system will classify the users in different classes based on their previous visits as different users have different sets of recommendations, thus when user again visits the site a personalised page of recommendation is provided to the user.

4.2. Application of Web Data Mining in E-Commerce

The emergence of E-commerce, changes the business philosophy. It has brought the massive change for various industries. To extract information and information management, the blend of data mining technologies and E-commerce can significantly improve the capability of the enterprise. The applications of web data mining in E-commerce are as follows [15]: Keep old customers and develop new ones, enterprise resource optimization, improve the site design and clustering the customers.

4.2.1. Keep old customers and develop new ones

With the business philosophy that "customer-focus" is now days getting support among people, to analyse customers and to understand their needs and on the basis of their needs to guide them is become an important topic for the enterprise management. With the help of data mining, E-commerce sites understands the users' needs and their preference, on basis of that site developers restructure their sites and provide personalised services to the customers so as to keep the old customers. Through the analysis of new customers and their access, mining these valuable customers and guiding them related to their needs and make then become a real user.

4.2.2. Enterprise resource optimization

For every enterprise or corporate cost saving is the key to earning. Through the data mining technologies by analysing of historical financial data, transactional data, and analysing of main input and output activities, real-time and accurate information is grasped by the enterprise. By this the enterprise can optimize the allocation of resources effectively. Through data mining the E-commerce enterprise can obtain reliable information to analyse the customers' behaviour and its spending pattern for future. So as to understand the customers

browsing pattern it is being decided by the enterprise the actual position of advertisement to increase the advertisement investment rate and reduce the operational costs results improvement in competitiveness of enterprise.

4.2.3. Improve the site design

Mining the customers visiting pattern and their behaviour and feedbacks, that is used by the developers to design the site according to the visitors' desire. By using the association rule in the web the developer can dynamically adjust the site structure for different customers according to their needs, so the customers can easily access the page what they want. This leaves a good impression on the customers and increases the probability of their next visit to the site. A well designed website is always measured as a successful website [11]. It must be designed such a way that it can always gratify the needs of the users.

4.2.4. Clustering of customers

Customer clustering is the main component in E-commerce. With the similar browsing behaviour of customers can be grouped and common characteristics of the similar grouped customers analysed so as to help the E-commerce to understand the common needs of the customers and meet the customers' needs in many ways. This makes the E-commerce to provide a high quality and good service to the customers.

5. Conclusion

The web data mining is a very fast developing area now days. Based on the analysis of the concepts and application of web data mining, this paper combines the web data mining technologies and e-service platform to help enterprise to find valuable knowledge in the network information resources. In E-commerce, it can predict the current trends of the customers as a result helping business to retain the old customers and finding the new one. By making correct decisions results improve in competitiveness of enterprise. Web data mining also plays an important role in E-learning. At present learners are highly dependent on web. As a large number of E-learning sites are being designed to provide useful information to the learners. In order to make a site popular among its visitors, site designer always try to increase its effectiveness by understanding users and learning. The web designers study the learners' requirements and their preference to maintain and restructure their sites. The proposed mining strategies and recommender agents may further be improved by employing innovative research areas like cloud computing and soft computing techniques.

References

- [1] A. Martin, C. León, "An intelligent e-learning scenario for knowledge retrieval". IEEE.
- [2] C.J. Carmona, S. Ramirez-Gallego, F. Torres, E. Bernal, M.J. del Jesus, S. Garcia, "Web usage mining to improve the design of an e-commerce website: OrOliveSur.com", Expert systems with applications 2012, Elsevier, doi:http://dx.doi.org/10.1016/j.eswa.2012.03.046.
- [3] Dr. Jamil A. Itmazi, "E-Learning Systems and Tools" 1st editing, Publisher: Phillips Publishing, on April 2010, ISBN: 978-0-557-44966-8.
- [4] Dr. S.S. Gautam, Manish Kumar Tiwari, "Web Mining-Concepts and its Applications", International Research Journal of Computer Science (IRJCS), Issue 01, Volume 3 (January 2016), ISSN: 2393-9842, pp. 8-13.
- [5] M. M. Shahabadi, M. Uplane, "Synchronous and asynchronous e-learning styles and academic performance of e-learners", Procedia-Social and Behavioral Sciences 176(2015), pp. 129-138.
- [6] Neha Verma, Prof. Jatinder Singh, "Improved Web Mining for E-commerce Website Restructuring" IEEE International Conference on Computational Intelligence & Communication Technology, 2105, doi:10.1109/CICT.2015.76, pp. 155-160.
- [7] Prajyoti Lopes, Bidisha Roy, "Dynamic Recommendation System Using Web Usage Mining for E-commerce Users", International Conference on Advance Computing Technologies and Applications (ICACTA-2015), Elsevier, doi:10.1016/j.procs.2015.03.086.
- [8] R. Cooley, "Web Usage Mining: Discovery and application of interesting patterns from web data", PhD thesis, Department of Computer Science, University of Minnesota, May 2000.
- [9] R. Cooley, B. Mobasher, J. Srivastava, "Data preparation for mining world wide web browsing patterns", Knowledge and Information Systems, 1(1), 1999.
- [10] Sunil, Prof. M.N. Doja, "A Review Paper on Identifying Students Interest in E-learning using Web Usage Mining", International Journal of Latest Trends in Engineering and Technology (IJLTET) Volume 8, Issue 1, pp.520-525, e-ISSN:2278-621X, doi:http://dx.doi.org/10.21172/1.81.067.
- [11] Sunil, Prof. M.N. Doja, "Recommender System Based On Web Usage Mining For Restructuring Of E-Learning Websites And Blogs", International Journal of Research in Engineering & Advanced Technology (IJREAT), Volume 5, Issue 1, Feb - Mar, 2017, ISSN: 2320 - 8791.
- [12] Xinjin Li, Sujing Zhang, "Application of web usage mining in e-learning platform", International Conference on E-business and E-government, IEEE 2010, doi:10.1109/ICEE.2010.353., pp. 1391-1394.
- [13] Yanduo Zhao, "The review of web mining in E-commerce", International Conference on Computational and Information Sciences, IEEE 2013, doi:10.1109/ICCIS.2013.158, pp.571-574.
- [14] Ying Jia, "Application of web usage mining in distance education", shanxi: shanxi normal university, 2008, Springer.
- [15] Yonghua Zhao, Hong Lin, "WEB data mining applications in e-commerce", The 9th International Conference on Computer Science & Education (ICCSSE 2014), August 22-24, IEEE 2014 Vancouver, Canada. pp. 557-559.