

A Survey on approach of Security in E-Commerce using Web Mining Framework

Afrin M Sheikh¹ Sumitra Menaria²

¹M.E Student ²Assistant Professor

^{1,2}Parul Institute of Engg. And Technology Parul University-Vadodara

Abstract— Web Data Mining is an important area of Data Mining which deals with the extraction of interesting knowledge from the World Wide Web, It can be classified into three different types i.e. web content mining, web structure mining and web usages mining. In this paper, we introduce a web data mining solution to e-commerce to provide better services and use encryption for security. This paper also reports the summary of various methods of web data mining.

Key words: Web mining techniques, E-commerce, Electronic Commerce, Data mining, Web mining

I. INTRODUCTION

The Web is a critical aspect of communication and promoting a company image. The research aim is to provide security and discover Web information from massive sources of data (I.e. web data) using various web mining techniques. The research on the Web deals with representations of semi-structured and heterogeneous data, such as textual information, hyperlink structure and receives information of use to improve the quality of services provided by different Web applications. Users face many problems due to the huge volume of information that is consistently growing. In particular, Web users have issues in getting the correct information due to low precision.

For example, if a user wants to get any information by using Google and other search engines, it will provide not only Web contents dealing with this topic, but a series of irrelevant information, so called noise pages, resulting in difficulties for users in obtaining necessary information.

Web Mining Framework System:

In general, Web mining tasks can be classified into three categories:

- 1) Web content mining,
- 2) Web structure mining and
- 3) Web usage mining[3].

Web Content Mining:

It deals with finding useful information or knowledge from the web page contents. Margaret H. Dunham [11] stated Web Content Mining can be thought of as extending the work performed by basic search engines. Web content mining analyses the content of various Web resources. Recent advances in multimedia data mining promise to widen access also to image, sound, video, etc. content of Web resources. The primary Web resources that are mined in Web content mining are individual pages. Information Retrieval is one of the research areas that provide a range of popular and effective, mostly statistical methods for Web content mining.

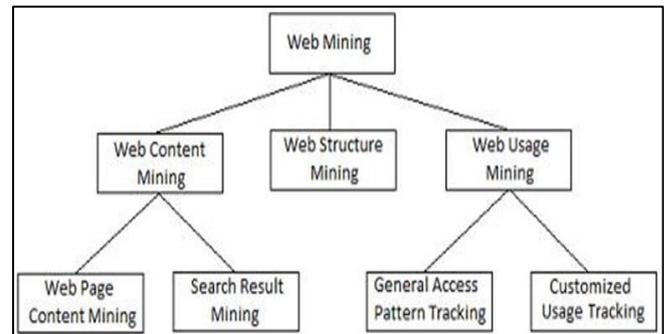


Fig. 1: Taxonomy of Web Mining

Since it is not possible to annotate images on internet manually. So Web images are usually not well annotated using semantic descriptors. Due to multiplicity of contents in a single image and the subjectivity of human perception, it is hard to make exactly the same understanding to the similar image by different users. These problems have restricted the application of the keyword based image retrieval tools and also traditional text based methods could not handle the explosive load of images and hence the concept of Content Based Image Retrieval was born. Content-Based Image Retrieval (CBIR) attempts to automate the process of indexing or annotating image in image databases.

Web Structure Mining (Web Linkage Mining):

Web structure mining has a number of challenges for dealing with the structure of the hyperlinks within the Web itself. Link analysis is very famous area of research. However, with the rising interest in Web mining, the research of structure analysis had improved and these hard works had resulted in a newly emerging research area known as Link Mining [17], which is located at the intersection of the work in link analysis, hypertext and web mining, relational learning and inductive logic programming, and graph mining. There is a potentially wide range of application areas for this new area of research, including Internet.

In the web structure mining we can have the following directions:

- Based on the hyperlinks, categorizing the Web pages and generated the information.
- Discovering the structure of Web document itself.
- Discovering the nature of the hierarchy or network of hyperlinks in the Website of a particular domain.

This phase analyses a web site by using both page rank algorithm and trust rank algorithm. The ranking of a page is determined by its link structure instead of its content. The trust rank algorithm is procedure to rate the quality of web sites. The output is quality based score which correspond to trust assessment level of the web site. The initial step is collects information from web sites and stores those web pages into web repository.

A. Page Rank Algorithm:

Page rank algorithm used by search engine. We have computed page rank of web sites by parse web pages for links, iteratively compute the page rank and sort the documents by page rank engine .Page Rank algorithm is in fact calculated as follows

$$\text{PAR}(A)=(1-d) +d(\text{PAR}(T1)/\text{OG}(T1) + \dots + \text{PAR}(TN)/\text{OG}(TN))$$

Where PAR(A) is the PageRank of page A

OG(T1) is the number of outgoing links from page T1

d is a damping factor in the range $0 < d < 1$,usually set to 0.85

The PageRank of web page is calculated as sum of the PageRank of all pages linking to its divided by the number of links on each of those pages its outgoing link[13].

Web Usage Mining:

Web usage mining is the most relevant part in terms of marketing because it explores ways to navigate and conduct during a visit to the website of a company. In web usage mining, there is analyzed information on web pages visits that are saved in log files of Internet servers in order to discover the previously unknown and potentially interesting useful patterns. Web usage mining is described as applying data mining techniques on Web access logs to optimize web site for user's interest. The continuous growth in the size and use of the Internet is creating difficulties in the search for information. To help users search for information and organize information layout, Smith and Ng [14] suggested using a self organizing map(SOM) to mine web data and provided a visual tool to assist user navigation. Based on the users' navigation behavior, they developed LOGSOM, a system that utilizes SOM to organize web pages into a two-dimensional map. The map provides a meaningful navigation tool and serves as a visual tool to better understand the structure of the web site and navigation behaviours of web users.

Most data used for mining [15] is collected from Web servers, clients, proxy servers, or server databases, all of them produce noisy data. Because Web mining is sensitive to noise, data cleaning methods are necessary. Markov models have been extensively used to model Web users' navigation behaviours on Web sites. Jianhan Zhu, Jun Hong et al[16] proposed a clustering algorithm called Citation Cluster to cluster conceptually related pages. The clustering results are used to construct a conceptual hierarchy of the Web site. Markov model based link prediction is integrated with the hierarchy to assist users' navigation on the Web site.

In the previous six years collection of user navigation session were presented in form of many models such as Hyper Text Probabilistic Grammar (HPG), N-Gram Model, Dynamic clustering based morkov model etc [17].

Web usage mining analyzes information about visited Web page that saved in log files of Internet servers in order to discover interesting patterns previously unknown and potentially useful. Web usage is described as mining applying data mining techniques on Web access logs to optimize web site for users.

II. SECURITY PERSPECTIVE

Lack of trust is one of the main reasons which can make e-commerce less attractive because of the fear of credit card number or sensitive information being stolen[12]. The increasing number of the web security attacks causes fears to

consumers that resulted in lack of trust. Hence, many businesses and internet users are reluctant to use the new technology.

According to the largest internet security company McAfee, almost half of consumers had terminated an order or due to security fears. Even in an attempt to get a good deal, 63% consumers will refuse to purchase from a Website that does not show a Trustmark or security policy[7]. Usually, e-commerce firms seek to get trust of their users by creating and advertising new security strategies, but the security threat is still growing and affecting e-commerce firms negatively.

A. E-Security Issues and Trust:

“A security threat has been known as a situation, or event with the potential to effect economic adversity to data or network resources in the form of destruction, disclosure, modification of data, denial of service, and/or fraud, waste, and abuse Security, then, is the protection against these threats”. Under this definition, threats can be made either through network and data transaction attacks, or via unauthorized access by means of defective authentication. For customers, it must be recognized that economic hardship encompasses damages to privacy as well as theft, of credit information and authentication issues for consumers will be overturned; as in whether the Web site is „real” rather than whether the purchaser's identity is real. This modified definition explains the security threats from a consumer's point of view. Security in B2C electronic commerce is reflected in the technologies used to secure customer data. Security concerns of consumers may be addressed by many of the same technology protections as those of businesses, such as encryption and authentication[1].

Because of all these security issues there is a great need of web security. Therefore the proposed system will implement the security by implementing the encryption ciphering techniques.

B. E-COMMERCE Security Tools:

- Firewalls – Software and Hardware
- Public Key infrastructure
- Encryption software
- Digital certificates
- Digital Signatures
- Biometrics – retinal scan, fingerprints, voice etc
- Passwords
- Locks and bars – network operations centers

Observation and Finding:

A survey was conducted on E-commerce privacy and security system so as to collect views of different people (student, professional and others). The objective of this survey is to obtain the knowledge of e-commerce privacy and security system prevalent among the users. The responses of people have been represented through graphs and tables on the basis of responses obtained from people. The graphs of questionnaire along with their explanation have been discussed below:-

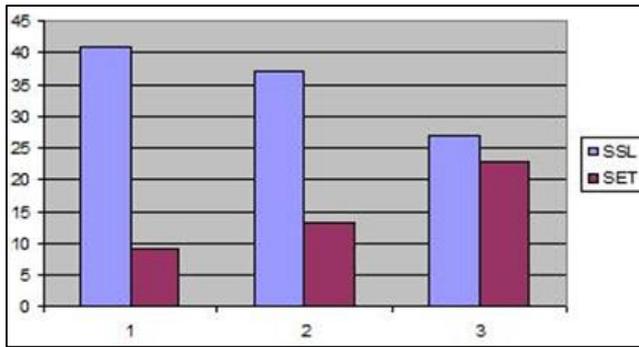


Fig. 2: Graph Represented Yes/No options:

S.No.	Questions	Yes	No
1	Is E-commerce security critical	40	10
2	Are Consumer communicating With correct server	38	12
3	Is privacy can achieved by using encryption decryption techniques	44	6
4	Do banks provide secure banking service	48	2

Table 1:

In this graph the minimum responses of people have been given to the question no. 3 “Can privacy can achieved by using encryption and decryption techniques?” 38 persons were affirmations and 12 persons were negations but in reality the Encryption /decryption techniques are play a great role to achieve privacy . This technique is the best way to take our information private.

III. OUR CONTRIBUTION

This paper proposes a system model which is very useful in e-commerce applications and its security. This system involves integration of web mining framework with an e-commerce application involving the clustering and classification techniques and also with Encryption technique for providing security.This integration facilitate e-store owner to improve the features and security.There are many areas where data mining can be very helpful when integrating with e-commerce.

Classification is one of the Data Mining techniques that is mainly used to analyze a given data set and takes each instance of it and assigns this instance to a particular class such that classification error will be least. It is used to extract models that accurately define important data classes within the given data set. Classification is a two step process. During first step the model is created by applying classification algorithm on training data set then in second step the extracted model is tested against a predefined test data set to measure the model trained performance and accuracy. So classification is the process to assign class label from data set whose class label is unknown.

Some existing clustering techniques: Hierarchical agglomerative Clustering is used to identify unique skill set clusters. The classification of product is validated into clusters by performing k-means cluster analysis[5].

A. Hierarchical Agglomerative Clustering:

Hierarchical clustering is a bottom-up clustering method where clusters have sub-clusters, which in turn have sub-

clusters, etc. Agglomerative hierarchical clustering starts with every single object (gene or sample) in a single cluster. Then, in each successive iteration, it agglomerates (merges) the closest pair of clusters by satisfying some similarity criteria, until all of the data is in one cluster. The hierarchy within the final cluster has the following properties: Clusters generated in early stages are nested in those generated in later stages. Clusters with different sizes in the tree can be valuable for discovery. A Matrix Tree Plot visually demonstrates the hierarchy within the final cluster, where each merger is represented by a binary tree[5].

B. K-Means Cluster Analysis:

K-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean[5]. It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data as well as in the iterative refinement approach employed by both algorithms.

Procedure: The input dataset is partitioned into K number of clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. For each data point: Calculate the distance from the data point to each cluster. If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster. Repeat the above step until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends. The choice of initial partition can greatly affect the final clusters that result, in terms of inter-cluster and intra cluster distances and cohesion.

IV. CONCLUSIONS

Web data mining is a fast rising research area today. As the web data and its usage will rise in future. It will prolong to generate more content, structure and usage data. So the importance of web data continues increasing. E-commerce security is the protection of e-commerce assets from unauthorized access, use, alteration, or destruction. In the proposed model web mining integrated with the electronic commerce application to improve the performance of e-commerce applications and provided with better security.We have conclude and understand some important taxonomy of web mining which are used in data mining. After that we get a view of the proposed techniques and flow which contains mainly user identification phase which involves user authentication task by applying encryption techniques .

REFERENCES

- [1] R.Manjusha, Research Scholar, Dr. R.Ramachandran, “Web Mining Framework for Security in E-commerce”, IEEE - International Conference on Recent Trends in Information Technology,ICRTIT 2011
- [2] Wang Yong gui,Jia Zhen, “Research on Semantic Web Mining” ,2010 International Conference On Computer Design And Applications (ICCCA 2010)The Internet and Management, Vol. 10, No.3, 2002, p 1 - 14
- [3] Jinhua sun,Yanqi Xie, “A Web Data Mining Framework for E-commerce Recommende Systems”, 978-1-4244-4507-3/09

- [4] SaiMing Au, "A Study Of Application Of Web Mining For E-Commerce: Tools And Methodology", International Journal of The Computer,
- [5] "Web Mining In E-Commerce" , Istrate Mihai University of Pitești Faculty of Mathematics and Informatics No.1,Stadionului, Țicleni, Gorj, 215600, Romaniamihaifrance@yahoo.com +40 745 775 935
- [6] <https://arxiv.org/ftp/arxiv/papers/1311/1311.7388.pdf>
- [7] YacineAtif , "Building Trust in Ecommerce",1089-7801/02 ©2002 IEEE.
- [8] <http://wwwusers.cs.umn.edu/~desikan/publications/wmo.pdf>
- [9] <http://www.ycjianguo.xyz/read-book/web-mining-applications-in-e-commerce-and-e-services.pdf>
- [10] https://www.researchgate.net/publication/220520203_WEBKDD_2000_-_Web_Mining_for_E-Commerce
- [11] Margaret H. Dunham, "Data Mining Introductory & Advanced Topics", Pearson Education
- [12] Ahmad TasnimSiddiqui , Arun Kumar Singh , "Secure E-business Transactions By Securing Web services"978-0-7695-4853-1/12 © 2012 IEEE.
- [13] Mrs. Pallavi Chandratre, Prof. Umesh Kulkarni "Implementation of trust ranking algorithm", International Journal of Science Technology & Management, Volume No.04., February 2015
- [14] K. A. Smith and A. Ng, Web page clustering using a self-organizing map of user navigation patterns, Decision Support Syst. 35(2) (2003) 245–256
- [15] Jaideep Srivastava, R. Cooley, "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data", ACM SIGKDD, VOL.7
- [16] Jianhan Zhu, Jun Hong et al," Using Markov Models for Web Site Link Prediction" College Park, Maryland, USA ACM June 11-15, 2002
- [17] Borges and M. Levene,"A dynamic clustering-based markov model for web usage Mining", cs.IR/0406032, 2014
- [18] Kavita Sharma, Gulshan Shrivastava, Vikas Kumar "Web Mining: Today and Tomorrow" – IEEE 2011
- [19] Ashish Jain, Rajeev Sharma, Gireesh Dixit and Varsha Tomar "Page Ranking Algorithms in Web Mining, Limitations of Existing methods and a New Method for Indexing Web Pages" - IEEE 2013
- [20] Niranjanamurthy , DR. Dharmendra Chahar "The study of E-Commerce Security Issues and Solutions" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 7, July 2013
- [21] Kuldeep Kaur Int. Journal of Engineering Research and Applications "E-Commerce Privacy and Security System" ,ISSN : 2248-9622, Vol. 5, Issue 5, (Part -6) May 2015, pp.63-73
- [22] Joseph P.T,S.J(2008),"An Indian perspective", 3rd edition, E-Commerce,by PHI learning private limited.
- [23] Kima S , " An Empirical Study of Customers perceptions of security and trust in e-payment system," Electronic Commerce research and applications, Vol .9 No. 1 pp.84-95(2010)
- [24] Corbitt,B.J. , " Trust and e-commerce a study of e-commerce perceptions", Electronic Commerce Research & Application, Vol .2 No.3 ,pp.203 -15