

Implementation of Trust Issues in Ecommerce

Mrs.Pallavi Chandratre

ME computer
Vidyalankar Institute of Technology,
Mumbai.
E-mail pallavi.chandratre@vit.edu.in

Prof. Umesh Kulkarni

DepartementOf Computer Engg.
Vidyalankar Institute of Technology,
Mumbai
E-mail umesh.kulkarni@vit.edu.in

Abstract--This paper is based on security of E-commerce web sites. It uses web mining technology for providing security on e-commerce web sites. The connection between web mining security and ecommerce analyzed based on user behavior on web . Different web mining algorithms and security algorithm are used to provide security on e-commerce web sites. Based on customer behavior web mining algorithms like page rank algorithm and trust rank algorithm is used for developing web mining framework in e-commerce web sites. This application will develop false hit database algorithm and nearest neighbor algorithm to provide security on e-commerce web site. It will also develop page rank & trust rank algorithm for providing security on e-commerce site.

Keyword–Web mining, Security, E- commerce.

I. INTRODUCTION

The World Wide Web popularity leads to a revolution towards electronic commerce. Network transactions, electronic payments and on-line receipts are changing the traditional ways of doing business. Many companies take benefits of the e-commerce chances and other institutions will follow. The rapid growth of e-commerce is attracting the attention of businesses with its characteristics high-efficiency, low cost, high-profitability and global application. However, Security fears cause million dollars loss for e-commerce retailers .

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[4]. 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. 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. The issues of available reliable security technology and exploitation are not only limited to e-commerce technologies, but also broadly impacting computer and information systems throughout the world especially in developing countries because there are many

gaps and lack of awareness as they are still at the exploratory stages.

i 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. This definition must be tailored in order to be appropriate to consumer transactions to acknowledge that consumer information has value. 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 Trust Rank algorithm. The Proposed System consist of three phase’s web structure mining analysis, Web Content Mining analysis, decision analysis.

II. WEBMINING FRAMEWORK SYSTEM

Web mining is the use of data mining techniques to automatically discover and extract knowledge from web documents. web mining is the information service centre for news, e-commerce, and advertisement, government, education, financial management, education, etc. We have developed Web mining framework for evaluating ecommerce web sites[1]. In general web mining task can be classified into web content mining, web structure mining and web usage mining. Some of the well-known classification techniques for web mining such as like, page rank algorithm and trust rank algorithm is used in this paper.

III. WEB STRUCTURE MINING ANALYSIS

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.

i 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

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

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.

ii TRUST RANK ALGORITHM

The trust rank algorithm is procedure to rate the quality of web sites. Taking the linking structure to generate a measure for quality of a page. Steps of Trust Rank algorithm.

- 1 The starting point of the algorithm is the selection of trusted web pages.
- 2 Trust can be transferred to other page by linking to them.
- 3 Trust is propagating in the same was as Page Rank
- 4 The negative measure is propagating backwards and is a measure of bad pages
- 5 For the ranking algorithm both measures can be taken into account.

Trust Rank algorithm is in fact calculated as follows

Trust Rank = Trust Rank Of the User + Trust Rank of the Web Page

Trust Rank Of The User : It uses nearest neighbor algorithm. Whenever user register into the system he has to give the reference of the previous trusted user. The trust value of the user is calculated as

$$\text{Trust Value} = \text{Trusted value of the referenced user} / 2$$

Trust rank of Web pages :

As the user visit the web page he has to assign the rank value to that particular page.

The number of user visited to that page will give the different rank values to that page. Therefore the final trust rank of the page is calculated as

Trust value of the Page = Average trust value assign by each user to the page

IV. WEB CONTENT MINING ANALYSIS

Web content mining is defined as searching of new information from web data. Data is retrieved for desired topic by user. In Web content mining analysis we have taken example product categories & their features associated with them.

We are performing a cluster analysis on the products in two phases. Hierarchical agglomerative

Clustering is the first step to identify unique skill set clusters. The classification of product is validated into clusters by performing k-means cluster analysis.

i HIERARCHICAL AGGLOMERATIVE CLUSTERING

Hierarchical clustering is a bottom-up clustering method where clusters have sub-clusters, which in turn have sub-clusters, etc. The classic example of this is species taxonomy. Gene expression data might also exhibit this hierarchical quality (e.g. neurotransmitter gene families). 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.

Process: Assign each object to a separate cluster. Evaluate all pair-wise distances between clusters. Construct a distance matrix using the distance values. Look for the pair of clusters with the shortest distance. Remove the pair from the matrix and merge them. Evaluate all distances from this new cluster to all other clusters, and update the matrix. Repeat until the distance matrix is reduced to a single element.

Advantages: It can produce an ordering of the objects, which may be informative for data display. Smaller clusters are generated, which may be helpful for discovery.

ii. 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. 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.

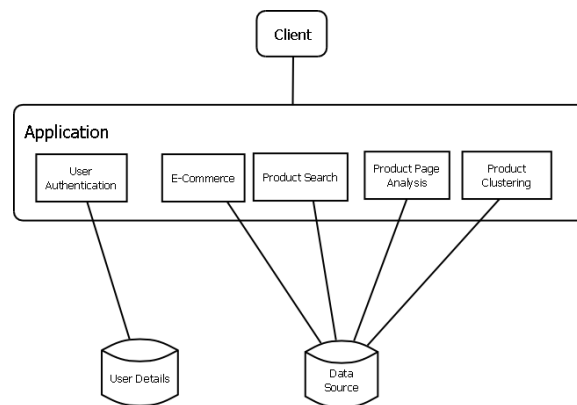
Process: The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points. For each data point: Calculate the distance from the data point to each cluster. If the data point is closest to its own 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.

Advantages: With a large number of variables, K-Means may be computationally faster than hierarchical clustering (if K is small). K-Means may produce tighter clusters than hierarchical clustering, especially if the clusters are globular.

V. DECISION ANALYSIS

This phase uses the total trust of web page generated from Web structure mining analysis phase to perform the Trust calculation of web site.

VI. MODULES OF THE PROPOSED SYSTEM & ITS ARCHITECTURE



Architecture of the proposed system consist of the following modules

i. USER IDENTIFICATION

Users are of different categories. New Users will get registered in the system. Existing users can logon to their account. Administrator has the highest priority. Generate user profiles based on their access patterns. The administration login module is as follows.

Admin Login

The user login module is as follows.

User Login

ii. E-COMMERCE MODULE

This will facilitate users to add products and their features into the system in a particular format. It will also facilitate to view the valid product list populated. This will

lead to generation of pages related to products which will be validated for its trust level based on the Trust Rank mechanism.

1) Add Product Module

This module facilitate the user to add the products & their features into the system

2) Edit Profile Module

User can edit his own profile by using this module.g.change address, change email-id etc.

3) Pending Approval Module

As each user registered in the system he has to give reference email address of another user. User can enter into the system if the refered user will approve him. This module will display the pending approval for the user.

Full Name	User Name	Gender	Email Address	Contact		
15	user2	user2	Male	user2@gmail.com	9969372495	Approve
15	user3	user3	Male	user3@gmail.com	9978451256	Approve

iii PRODUCT SEARCH MODULE

This module will facilitate searching of the products present in the product list based on the keyword specified as well as the Trust Rank associated with the product. Thereby, the user will be facilitated to perform an accurate searching of products.

As the user enter the product to be search it will display all the products available in the system according to the keyword & the features of the product.

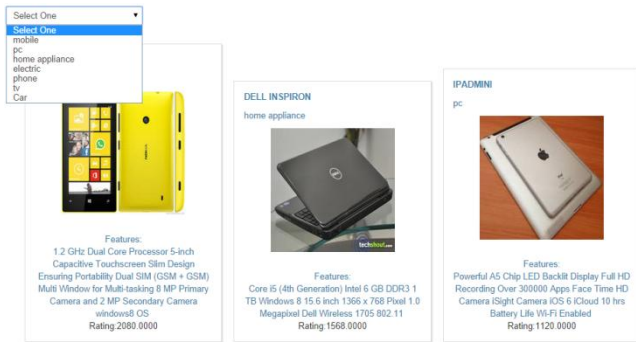
iv PRODUCT CLUSTERING

This module will facilitate the system to categorize the products into various categories based on a selected set of characteristics. These clustered products will be utilized for performing the search process. The cluster database are as shown

id	cluster_name	cluster_values	createdtime
1	Secondary 4D Primary and GHz Capacitive Processor...	1.2,3.4,35,35	2014-09-18 11:25:04
2	8 GB Core	5.6,18,21,32	2014-09-18 11:25:04
3	Camera 3 ISO	7.8,9,13,19,22,23,26,31,34	2014-09-18 11:25:04
4	Camera	10,14,15,16	2014-09-18 11:25:04
5		12,12	2014-09-18 11:25:04
6		25	2014-09-18 11:25:04
7		27	2014-09-18 11:25:04
8		29,30	2014-09-18 11:25:04

v PRODUCT PAGE ANALYSIS

This module will facilitate the system to analyze various product pages and finalize the Trust Rank associated with them. Thereby, the product validity will be decided by the system based on the Rank associated with them. Thus, only valid products will be displayed by the application on the product page.



VII. CONCLUSION

In this paper we have proposed web mining framework for e-commerce web sites. In web mining framework we have developed three phases' web structure mining analysis, Web Content Mining analysis, decision analysis. In web structure mining analysis we have used page rank algorithm and trust rank algorithm. In Web Content Mining analysis we have used Hierarchical agglomerative clustering and k-means cluster analysis. In decision analysis we have used trust calculation of web site to analyses the result of the evaluation.

REFERENCES

- [1] R. Manjusha, Dr.R.Ramachandran, "Web Mining Framework for Security in E-commerce", IEEE-ICRTIT 2011.
- [2] Chuck Litecky, Andrew Aken, Altaf Ahmad, and H. James Nelson, Southern Illinois University, Carbondale, Mining computing jobs, January/February 2010 IEEE SOFTWARE.
- [3] Lu Tao & Lei Xue, "Study on Security Framework in E-Commerce", 1-4244-1312-5/07 © 2007 IEEE.
- [4] Ahmad Tasnim Siddiqui, Arun Kumar Singh, "Secure E-business Transactions By Securing Web services" 978-0-7695-4853-1/12 © 2012 IEEE.
- [5] Jun Li, Dongting Yu, Luke Maurer, "A Resource Management Approach to Web Browser Security" 978-1-4673-0009-4/12 © 2012 IEEE.
- [6] Sankar K. Pal, Fellow, IEEE, Varun Talwar, Student Member, IEEE, and Pabitra Mitra, Student Member, IEEE, "Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 5, SEPTEMBER 2002.
- [7] Yacine Atif, "Building Trust in E-commerce", 1089-7801/02 © 2002 IEEE.
- [8] Hatoon Matbouli & Qigang Gao, "An Overview on Web Security Threats and Impact to E-Commerce Success", 978-1-4673-1166-3/12 © 2012 IEEE.
- [9] Jun Li & Dongting Yu, Luke Maurer, "A Resource Management Approach to Web Browser Security", 978-1-4673-0009-4/12 © 2012 IEEE.