

Recommending Based on Analysis of User's Behaviour in An E-Commerce Website

Siva Jyothi Barla^{1*}, G.V. Gayathri^{2*}

¹Department of CSE, Anil Neerukonda Institute of Technology & Sciences, Sangivalasa, Visakhapatnam, India

²Department of CSE, Anil Neerukonda Institute of Technology & Sciences, Sangivalasa, Visakhapatnam, India

Available online at: www.ijcseonline.org

Accepted: 17/May/2018, Published: 31/May/2018

Abstract— The web mining is a process which plays a important role in analyzing the web users behaviour. The web usage mining is a part of web mining has a major influence on web personalization. The aim of this study is the taking out necessary information from web access log files and applying data mining technique for analyzing the shopping behaviour of the customers through online.

Keywords— Data Mining , E-Commerce , WebLog Analysis, Behavioural Patterns.

I. INTRODUCTION

In today's ever connected world, the way people shop has changed. People are buying more and more over the Internet instead of going traditional shopping. E-commerce provides customers with the opportunity of browsing endless product catalogues, comparing prices, being continuously informed, creating wish list and enjoying a better service based on their individual interests [1]-[4]. The growing electronic market is highly aggressive, featuring the possibility for a customer to easily move from one e-commerce when their necessities are not satisfied. As a consequence, e-commerce business analysts require to know and understand consumers' behaviour when those navigate through the website, as well as try to identify the reasons that motivated them to purchase, or not, a product . This will be useful in making some improvements in the website design with the aim of increasing its efficiency. Getting this behavioural knowledge will allow e-commerce websites to deliver a more personalized service to customers, to retain customers and to increase benefits.

However, discovering customer' behaviour and the reasons that guide their buying process is a very complex task. E-commerce websites provide customers with a wide variety of navigational options and actions: users can freely move through different product categories, follow multiple navigational paths to visit a specific product, or use different mechanisms to buy products, for example. Usually, these user activities are recorded in the web server logs[5]. Web server logs store, in an ordered way, the sequence of web events generated by each user. The very valuable users' behaviour is hidden in these logs, which must be discovered and analysed. A correct analysis can be subsequently used to

improve the website contents and structure, to adapt and personalize contents, to recommend products, or to understand the interest of users in specific products, for instance.

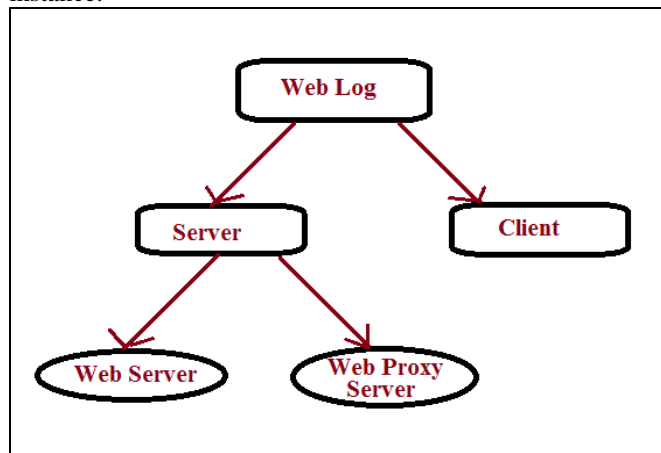


Figure 1. Represents the Basic Architecture of a Web Log

Data mining techniques have proved their usefulness for discovering patterns in log files (when applied to the analysis of web server logs the term *web usage mining* is used) which is shown in above figure 1. Its main goal is to discover usage patterns trying to explain the users' interests. Different techniques have been successfully used in the field of e-commerce, such as classification techniques, clustering, association rules or sequential patterns[6]-[9]. In many application domains these techniques are used in conjunction with process mining techniques. Such techniques are part of the business intelligence domain and apply specific

algorithms to discover hidden patterns and relationships in large data sets.

The initial step of web usage mining analysis is data pre-processing. The raw data have relatively low business value unless they can be transformed and processed to produce actionable knowledge. Therefore, in order to enable the analysis, raw logs must be pre-processed to discard uninteresting requests, to identify user sessions and to prepare the log to enable its analysis.

II. RELATED WORK

In this section we mainly describe about the related or back ground work that is carried out in order to propose this current application.

PREPROCESSING

The pre-processing step can be split into two main phases-

I. LOG CLEANING

The objective of this phase is to remove undesired records that may distort the results of the analysis.

II. USER IDENTIFICATION AND SESSIONIZATION

The aim of this phase is to group the events belonging to the same session (in terms of process mining, we are establishing the traces of processes).

Next, the process to be carried out is to analyze and identify behavioural patterns[10]. The goal of this analysis is to identify meaningful usage patterns that could be used to improve any website's design. In this percept, we are mainly interested in detecting the most relevant and used parts of the website and the relationship existing between them, as well as identifying behavioural patterns related to the buying process[11]-[14].

The process is as follows:

(a) Usage and Navigational Patterns: Next, we are going to explore some navigational patterns that illustrate users' preferences when browsing the website. Specifically, we analyse if sections are visited as the first option and if they are visited in an exclusive way.

(b) Behavioural Patterns Related to the Buying Process: Regarding the buying process there are two specifications that we are interested in- First, user sessions showing interest in acquiring a specific product that corresponds to the events of adding a product to the cart and adding a product to the wish list. In this regard, it is important to identify the sections visited just before such events happen. This way we could identify those sections that help users to find interesting products allowing correlating such information with different

access patterns. Second, sessions that buys some products, that is, sessions where the event Buy products in the cart happen. In this regard, it is important to analyse the relation between showing interest in a product and purchasing it.

III. METHODOLOGY

In this section we mainly describe about the methodology or algorithms that were applied for finding the Recommending based on Analysis of User's Behaviour in an E-commerce Website.

The following algorithms are applied in order to achieve the desired result.

- 1) Sequence Tree Algorithm
- 2) Apriori Algorithm

These two are the main algorithms that were used in this proposed application. Now let us look about this in detail as follows:

SEQUENCE TREE ALGORITHM

Sequential tree algorithm is used to find the frequent sequences in the log file. It involves two stages- construction of the sequence tree and mining of the sequence tree for finding frequent sequences. It takes as input the sequential database records which specify the various sequences of pages visited by the user and also threshold [15]-[18]. The frequent sequences that have been identified from the log file based on the minimum support threshold are obtained as output.

The sequence tree algorithm is described from steps (1) to (6).

Step 1:

Read the sequential database and create a map with key-value pairs. Key refers to the unique page name that was visited and sequence value refers to the frequency or number of times the page occurs.

Step 2:

Sort the map in descending order of frequency of the keys.

Step 3:

Create a sequence tree using the following steps

A root node is termed as null

For every row of sequential database that is read

Attach the sequence as a branch to the root.

If the sequence is already present, increment the counter
else a new branch is created with that sequence

Step 4:

Include a header table with number of rows equal to sequences with value one. Each row of header table is a linked list of nodes which specify the position where nodes are present.

Step 5:

Perform mining process

For every row in the header table

If frequency < minimum support threshold

Ignore the row of header table

else

For every node in the particular linked list

If frequency < minimum support threshold

Ignore the path

else

Traverse the tree up till the root and store the path in the file as frequent sequences. EndFor; EndFor

Step 6:**Output:**

The use of this algorithm results in finding out the most frequent users.

APRIORI ALGORITHM

It is a seminal algorithm, which uses an iterative approach known as a level wise search, where k-itemsets are used to explore (k+1)-itemsets.

It uses the Apriori property to reduce the search space: All nonempty subsets of a frequent itemset must also be frequent.

Procedure Apriori (T, minSupport)

```
{ L1= {frequent items};
```

```
for (k= 2; Lk-1 !=  $\emptyset$  ; k++)
```

```
{
```

Ck= candidates generated from Lk-1 for each transaction t in database

```
do
```

```
{
```

```
#increment the count of all candidates in Ck that are contained in t
```

```
Lk = candidates in Ck with minSupport } } U Lk ;
```

```
}
```

IV. RESULTS AND DISCUSSION

In this section we mainly discuss about the results and some other discussion related this proposed application.

Results Visualization and Interpretation

Finally, a graphical representation can be used to present raw results in a more simple, intuitive and useful manner. The use of this algorithm results in producing the most frequent item sets and the association rules based on them.

Notification

Once we have found out the most frequent users, the next step is to find out the most frequent items viewed by them. After finding out this we need to try to retain the customers by giving exciting deals about the products from the site owner's point of view.

This is done by adding a notification button which gives details about the discounts on the most frequently viewed or bought items.

Recommendation

Based on these rules, recommendations are given to the associated customers. This helps in increasing the sales from the owner's point of view as he is trying to attract the customers by introducing them to some other products which they might be interested to buy[19]. Thus, getting this behavioural knowledge will allow e-commerce websites to deliver a more personalized service to customers, to retain customers and to increase benefits [20].

Experimental Setup

In order to propose this current application, we used PHP as scripting and programming language with MY-SQL as the back end database to store the data set values. Here we took mysql tables for storing the data set inputs and then apply the above two algorithms in order to generate the processed output.

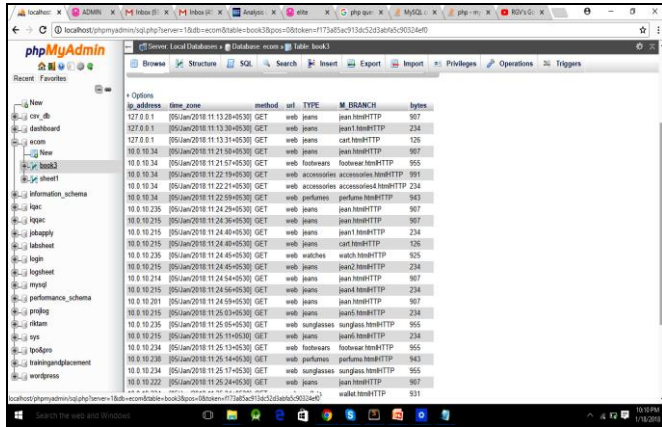


Figure 2.Represents the Data Set Information for the proposed Application

From the above figure 2 we can identify ECOM is the database name with three main tables like Book3, Sheet1 and New. So all these 3 tables are mainly used for storing the information into the database tables and now we can all the products information, accessories information in the book 3 table. If we want to add any new item or product into the dataset that information will be automatically added into the Book 3 Database only.

Sample Window for Sequence Tree Algorithm

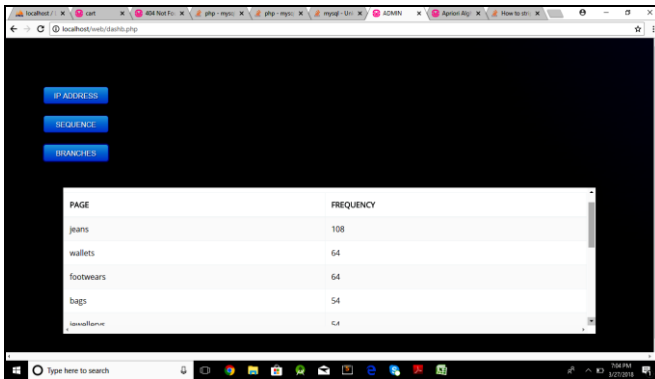


Figure 3.Represents the Output window for STA

From the above figure 3 ,we can clearly identify that there is a list of items displayed along with their frequency values in terms of integer values.Here for each and every product there will be a set of recommendations made by the individual users ,so all these recommendations are collected and finally their performance value is calculated by applying the STA approach.

Sample Window for Apriori Algorithm

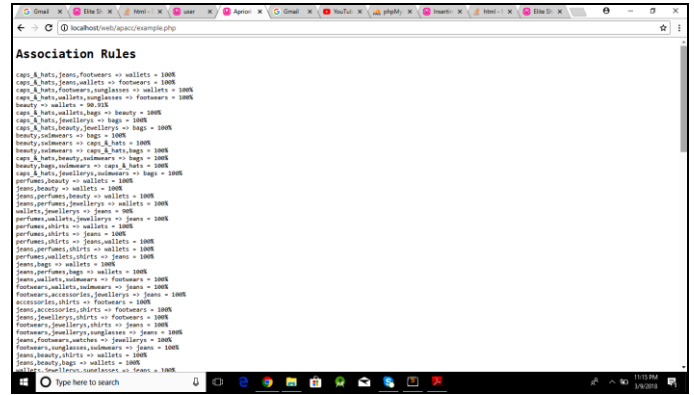


Figure 4.Represents the Output window for Apriori Algorithm

From the above figure 4 ,we can clearly identify that there is a list of items displayed along with their frequency values in terms of integer values.

V. CONCLUSION

In this application we finally proposed a novel method to extract or take out necessary information from web access log files and applying data mining technique for analyzing the shopping behaviour of the customers through online. By applying the proposed two algorithms we can able to extract the user recommendations for an appropriate product based on their individual area of interest. As we are using Apriori algorithm as core algorithm this will separate the individual products or categories into three categories like most rated items, average rated items and un rated items. This Apriori will try to find out the reviews based on a ranked manner and priority wise. By conducting various experiments on our proposed application, we finally came to a conclusion that this proposed method is best in providing accurate information like recommendations of products or items based on user ratings.

VI. REFERENCES

- [1] . Analysis of users’ behavior in structured e-commerce websites <http://ieeexplore.ieee.org/document/7933069/>
- [2]. Title-Sequential pattern mining from web log data http://www.ijesat.org/Volumes/2012_Vol_02_Iss_02/IJESAT_2012_02_02_10.pdf .
- [3]. Cooley, R., Mobasher, B., and Srivastava, J, “Web mining: information and pattern discovery on the World Wide Web”, International Conference on Tools with Artificial Intelligence, Newport Beach, IEEE,1997, pp. 558-567.
- [4]. Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava,” Data preparation for mining World Wide Web browsing patterns”, Journal of Knowledge and Information System,1999,pp. 1-27.
- [5]. Etzioni, The world-wide Web: quagmire or gold mine? Communications of the ACM 39 (11) (1996) 65–68.
- [6]. Chen Hu, Xuli Zong, Chung-wei Lee and Jyh-haw Yeh, “World Wide Web Usage Mining Systems and Technologies”, Systemic, Cybernetics and Informatics Vol. 1 – Number 4.
- [7]. Velayathan, Ganesan; Yamada, Seiji (2006). Behavior based web page evaluation. Proceedings of the 16th international conference on

- World Wide Web, 1317 – 1318. Retrieved May 13, 2010, from ACM Digital Library, <http://www.acm.org/dl>.
- [8]. Dilip Singh Sisodia and S.Verma “Application of Weblogs to Construct Smart Web servers to handle user Traffic efficiently’ International Journal of Advanced Computer Engineering and Architecture (IJACEA).
- [9]. J. Nielsen, “ Designing Web usability: the practice of simplicity. Indianapolis “ IN: New Riders Press, 2000
- [10]. A. Cooper, “The inmates are running the asylum. Indianapolis”, IN: SAMS, 1999.
- [11]. J. Preece, Y. Rogers and H. Sharp, “Interaction design”, New York, NY: John Wiley and Sons, Inc, 2002.
- [12]. M. Graff, “Individual differences in hypertext browsing strategies”, Behaviour and Information Technology, Vol. 24, no. 2, 2005.
- [13]. P. Pirolli, and S. Card, “Information foraging. Psychological Review”, Vol. 106, no. 4, 1999.
- [14]. L.D. Catledge, and J.E. Pitkow, “Characterizing browsing strategies in the World-Wide Web”, Computer Networks and ISDN Systems, Vol. 27, no.6, 1995.
- [15]. J. Holsanova, “ Tracking multimodal interaction with new media”, Paper presented at the workshop on The Citizen's Use and Comprehension of Information on the Internet, Uppsala, 2004..
- [16]. M.J. Bates, “The design of browsing and berrypicking techniques for the online search interface”, Online Review, Vol. 13, no. 5, 1989
- [17]. T. Bray, C.M. Sperberg-McQueen, “ Extensible markup language (XML) 1.0–W3C recommendation”, Technical Report REC-xml-19980210, World Wide Web Consortium, 1998.
- [18]. J.R. Punin, M.S. Krishnamoorthy, M.J. Zaki, “Web usage mining–languages and algorithms”, Technical Report, Rensselaer Polytechnic Institute, 2001.
- [19]. R. Gobinath and M. Hemalatha, “Optimized Feature Extraction for Identifying user Behavior in Web Mining ”, European Journal of Scientific Research, Vol. 105, no. 3, 2012.
- [20]. R. Gobinath and M. Hemalatha, “Improved Preprocessing Techniques for Analyzing Patterns in Web Personalization Process”, International Journal of Computer Application, Vol. 58, no. 3, 2013.