

# Study of user's behaviour in Structured E-Commerce Websites

**M.Tech.Scholar B.Thejaswini**

Dept. of Computer Science  
MJR College Engg.&Tech.  
Piler, Andhra Pradesh, India

**Asst.Prof.Reddi Durga Sree**

Dept. of Computer Science  
MJR College Engg.&Tech.  
Piler, Andhra Pradesh, India

**Karamala Suresh**

Dept. of Computer Science  
MJR College Engg.&Tech.  
Piler, Andhra Pradesh, India

**Abstract** -Online shopping is becoming more and more common in our daily lives. Understanding users' interests and behavior is essential in order to adapt e-commerce websites to customers' requirements. The information about users' behavior is stored in the web server logs. Nowadays, the growth of World Wide Web has exceeded a lot with more expectations. The internet is growing day by day, so online users are also rising. The interesting information for knowledge of extracting from such huge data demands for new logic and the new method. Every user spends their most of the time on the internet and their behaviour is different from one and another. Web usage mining is the category of web mining that helps in automatically discovering user access pattern. Web usage mining is leading research area in Web Mining concerned about the web user's behaviour. In this paper emphasizes is given on the user behaviours using web server log file prediction using web server log record, click streams record and user information. Users using web pages, frequently visited hyperlinks, frequently accessed web pages, links are stored in web server log files. A Web log along with the individuality of the user captures their browsing behaviour on a website and discussing regarding the behavior from analysis of different algorithms and different methods.

**Keywords** - Data mining, e-commerce, web logs analysis, behavioral patterns, model checking.

## I. INTRODUCTION

Nowadays the way people shop is totally different than the traditional way. People are buying more and more product online instead of going to the classical shop to shop to buy the product. E-commerce gives the opportunity to browse the number of different product with a different category, comparing different prices of products, create a wish list of product etc. e-commerce business is very competitive if the user does not get one thing at any site they can easily switch to another site for better options.

Therefore it is necessary to analyze the user's behaviour by the business analyst to give the better option and to motivate the user to buy the product. On the other hand study of user's behaviour on e-commerce sites is not an easy task. As this kind of application provides different navigation paths, users can navigate freely through the different category to a particular product. Generally, these users behaviour are stored in web server log, where it contains the ordered way or the sequence of user's activity created by users.

This log file is analyzed by an analyst to determine user's complex behaviour to increases the application contents and to provide proper suggestion to the user for the particular product. Generally, data mining algorithms are used to study these web server log files. The main approach of this kind of algorithms is to identify users

behavior and to find customers interest. Numbers of algorithms are proposed in recent years for data mining in the field of ecommerce such as classification techniques, clustering, association rules or sequential patterns. Their techniques are used along with data mining to discover hidden patterns and relationships in large datasets. Most of the data mining techniques used now days have some limitation in point of view to data mining for an e-commerce application.

They do not mine in the correct or proper sequence of the user's navigation sequence, they ignore causality relations such as users sequence, number of pages visited, product search sequence, number of time page visited by customer etc.

To limit all condition we proposed the use of Temporal Logic and model checking techniques as an alternative to the data mining technique. The main approach is to analyze users' behaviour on e-commerce site to discover customers' complex behavioural patterns by means of checking temporal logic formulas describing such behaviours against the log model.

At the start using web server log user behaviour is generated. After generation business analyst can use set of predefined queries which help him to discover the way client use the website.

## II. EXISTING SYSTEM

An e-commerce website is an open system where almost any customer behaviour is possible. This flexibility makes the discovery of a process-oriented model representing customers' behaviour a difficult task. This is so because there are so many different possible interactions that the final process model is either an over fitting spaghetti model or an under fitting flower model, from which no useful analysis can be done.

As a consequence, data mining techniques have been preferred for the analysis of e-commerce websites. Nevertheless, today's data mining techniques and tools have some constrains from the analysis point of view. On the one hand, they do not work in a direct way with the sequences of events (the click stream and all the data associated to each click) generated during the user's navigation through the website, but with an abstraction of such sequence, a kind of global photograph that ignores causality relations.

Such abstraction describes what happened during the session of a customer by means of a set of summarized data, such as the number of visited web pages, the frequency with which each product category was visited, or the time customers spend on a web page or category, for instance. On the other hand, most techniques are only able to classify these abstractions or discover simple relationships among certain high-level events of interest.

## III. DISADVANTAGES OF EXISTING SYSTEM

- For implementing mining techniques to predict the user's behaviour is computationally very expensive
- Currently, there are powerful commercial tools for analyzing logs of e-commerce websites, being Google Analytics one of the main ones.

They have some disadvantages with respect to the log-based analysis, such as dependence on JavaScript and cookies, the necessity of adding page tags to every page, the complexity of tag-based implementations, and the fact that, as a result, customers may experience a change in the download time of the website, or privacy concerns, for instance.

## III PROPOSED SYSTEM

In this paper we propose the use of Temporal Logic and model checking techniques as an alternative to data mining techniques. Such techniques have proved their applicability for open systems. We propose here a methodology for using it in structured e-commerce websites. The goal is to analyse the usage of e-commerce websites and to discover customers' complex behavioural patterns by means of checking temporal

logic formulas describing such behaviours against the log model.

At the beginning, web server logs are pre-processed to extract the detailed traces (sequences of events of a user session). Events can be user or system actions performed when a client visits a product or product category page, when he or she adds a product to the wishlist, when the search engine is used, etc. The business analyst can use a set of (predefined) temporal logic patterns to formulate queries that could help him to discover and understand the way clients use the website.

Considering the website structure and contents as well as the different types of user's actions, these queries can check the existence of complex causality relationships between events contained in the client sessions. From the tool point of view, the necessity of having control on the way the checking algorithms are applied, as well as the disappointing performance results we obtained when using some model checking tools at our disposal, mainly when used against big models, drove us towards the interest of developing a specific model checking tool. We did it using the SPOT libraries for LTL model checking.

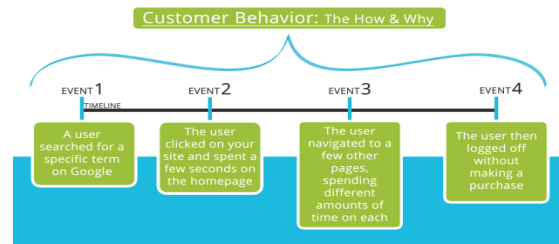


Fig.1 Behavior of the customer.

## IV. ADVANTAGES OF PROPOSED SYSTEM

- The advantage of our mining technique is that this provides causal relations among events of a user trace, instead of providing with a global view of the whole session.
- The proposed system can be executed in parallel, deploying different parallel servers with different parts of the log and executing the queries in parallel

## V. DATA PREPROCESSING

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.

The pre-processing step can be split into three main phases:

- Log cleaning.

- User identification.
- Log preparation.

## VI. IDENTIFYING USERS BEHAVIOURAL PATTERNS

The goal of this analysis is to identify meaningful usage patterns that could be used to improve the Up & Scrap website design. In this respect, 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 behavioral patterns related to the buying process. For that, we present the queries along with the question that they intend to answer, the results of the queries, the interpretation of the results and the actions proposed to improve the website, if any.

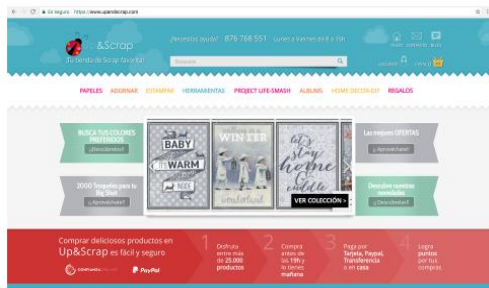


Fig. 2 Homepage of the Up and Scrap website.

## IV. CONCLUSION

In the case of open systems, where the sequences of interactions (stored as system logs) are not constrained by a workflow, process mining techniques whose objective is to extract a process model will usually provide with either over fitting spaghetti models or under fitting flower models, from which little interesting information can be extracted. A more flexible approach is required. In the paper we apply LTL-based model checking techniques to analyses e-commerce web logs. To enable this analysis, we have proposed a common way of representing event types and attributes considering the e-commerce web structure, the product categorization and the possibilities of users to navigate through the website according to such organization.

## REFERENCES

- [1]. J. B. Schafer, J. A. Konstan, and J. Riedl, "E-commerce recommendation applications." Hingham, MA, USA: Kluwer Academic Publishers, Jan.2001, vol. 5, no. 1-2, pp. 115–153.
- [2]. N. Poggi, D. Carrera, R. Gavalda, J. Torres, and E. Ayguad'e, "Characterization of workload and resource consumption for an online travel and booking site," in Workload Characterization (IISWC), 2010 IEEE International Symposium on. IEEE, 2010, pp. 1–10.
- [3]. R. Kohavi, "Mining e-commerce data: the good, the bad, and the ugly," in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2001, pp. 8–13.
- [4]. W. W. Moe and P. S. Fader, "Dynamic conversion behavior at e-commerce sites," *Management Science*, vol. 50, no. 3, pp. 326–335, 2004.
- [5]. G. Liu, T. T. Nguyen, G. Zhao, W. Zha, J. Yang, J. Cao, M. Wu, P. Zhao, and W. Chen, "Repeat buyer prediction for e-commerce," in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 155–164.
- [6]. J. D. Xu, "Retaining customers by utilizing technology-facilitated chat: Mitigating website anxiety and task complexity," *Information & Management*, vol. 53, no. 5, pp. 554 – 569, 2016.
- [7]. Y. S. Kim and B.-J. Yum, "Recommender system based on click stream data using association rule mining," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 320–13 327, 2011.
- [8]. R. Kosala and H. Blockeel, "Web mining research: A survey," *SIGKDD Explor. Newsl.*, vol. 2, no. 1, pp. 1–15, Jun. 2000.
- [9]. F. M. Facca and P. L. Lanzi, "Mining interesting knowledge from weblogs: a survey," *Data & Knowledge Engineering*, vol. 53, no. 3, pp. 225–241, 2005.
- [10]. C. J. Carmona, S. Ram'irez-Gallego, F. Torres, E. Bernal, M. J. Del Jes'us, and S. Garc'ia, "Web usage mining to improve the design of an e-commerce website: Orolivesur. com," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11 243–11 249, 2012.
- [11]. Q. Song and M. Shepperd, "Mining web browsing patterns for e-commerce," *Computers in Industry*, vol. 57, no. 7, pp. 622–630, 2006.
- [12]. O. Arbelaitz, I. Gurrutxaga, A. Lojo, J. Muguerza, J. M. Prez, and I. Perona, "Web usage and content mining to extract knowledge for modelling the users of the bidasoa turismo website and to adapt it." *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7478–7491, 2013.
- [13]. J. K. Gerrikagoitia, I. Castander, F. Reb'on, and A. Alzua-Sorzabal, "New trends of intelligent e-marketing based on web mining for e-shops," *Procedia-Social and Behavioral Sciences*, vol. 175, pp. 75–83, 2015.
- [14]. Y. H. Cho and J. K. Kim, "Application of web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, vol. 26, no. 2, pp. 233 – 246, 2004. K.-J. Kim and H. Ahn, "A recommender system using fGAg kmeans clustering in an online shopping market," *Expert*

Systems with Applications, vol. 34, no. 2, pp. 1200  
– 1209, 2008.

**Author's Profile**



**B. Thejaswini** Pursuing M.Tech at MJR College of Engineering & Technology, Department of CSE, Piler, Chittoor Dist.



**Reddi Durga Sree** Working as a Assistant Professor in MJR College of engineering & technology, Department Of CSE, Piler, Andhra Pradesh.



**Karamala Suresh** Working as a Head of the Department in MJR College of Engineering and Technology, Department Of Cse, Piler, Chittoor dist. A.P. India He is having 14 years of teaching experience in engineering colleges, he received B.Tech (CSE) from JNTU Hyderabad in 2002, Received M.E(CSE) from satyabama university Chennai in 2006, and Received M.Tech(CSE) in 2015 from JNTU Anantapuramu, he published several research papers in various national and international journals. He is interested in Computer Networks