Analysis of User's Behavior I Structured E-Commerce Websites

SAIDAM HARITHA[1], PUDOTA SRI VISMITHA[2], SANKA KAVYANJALI[3], SHAIK SHIREEN[4], MR.K.MOHAN KRISHNA[5]

[1,2,3,4,5] *Computer Science & Engineering, Vasireddy Venkatadri Institute of Technology, Guntur*

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. The analysis of such information has focused on applying data mining techniques where a rather static characterization is used to model users' behavior and the sequence of the actions performed by them is not usually considered. Therefore, incorporating a view of the process followed by users during a session can be of great interest to identify more complex behavioral patterns. To address this issue, this paper proposes a linear-temporal logic model checking approach for the analysis of structured e-commerce web logs. By defining a common way of mapping log records according to the e-commerce structure, web logs can be easily converted into event logs where the behavior of users is captured. Then, different predefined queries can be performed to identify different behavioral patterns that consider the different actions performed by a user during a session. Finally, the usefulness of the proposed approach has been studied by applying it to a real case study of a Spanish e-commerce website. The results have identified interesting findings that have made possible to propose some improvements in the website design with the aim of increasing its efficiency.

Index Terms: Data mining, e-commerce, web logs analysis, behavioral patterns, model checking.

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. This increasing electronic market is highly competitive, 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' behavior when those navigate through the website, as well as trying to identify the reasons that motivated them to purchase, or not, a product. Getting this behavioral knowledge will allow e-commerce websites to deliver a more personalized service to customers, retaining customers and increasing benefits. However, discovering customer' behavior and the reasons that guide their buying process is a very complex task

II. IDENTIFICATION AND RESEARCH.

An e-commerce website is an open system where almost any customer behavior is possible. This flexibility makes the discovery of a process-oriented model representing customers' behavior 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. A product review that conveys a high level of trustworthiness might not convey as much expertise, and vice versa.

For example, a long, detailed review of a laptop might signal a sincere and honest attempt to help others make

a good purchasing decision, but the review might contain accuracy problems that would lower the extent to which the review conveys expertise.
Professional communicators charged with managing user-generated content need not spend resources on channeling it into retailer and other independent review site environments as opposed to brand site environments.

III. PROPOSED SYSTEM AND ADVANTAGES

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 analyze the usage of e-commerce websites and to discover customers' complex behavioral patterns by means of checking temporal logic formulas describing such behaviors against the log model. At the beginning, web server logs are preprocessed 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 wish list, 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.

Millennial spend more than five hours per day with peer-created media. Millennial find user-generated content 35% more memorable than other media (including traditional media such as television).Millennial trust user-generated content50% more than other media. Millennial find user-generated content20% more influential on their purchase.

Modules: As professional communicators know, no communication occurs in a vacuum; all communication has context, an environment. This environment can critically influence how receivers interpret a message and whether they decide to act on it. Online product reviews—a type of user-generated content—appears in several environments: personal blogs such as Computer audio file.

1) Admin add product
2) Owner view product
3) User buy product

ADMIN ADD PRODUCT: Admin can login their account. Add the main category and next add the sub category. View the added product, if u wants to delete the files means delete.

OWNER VIEW PRODUCT: View the admin added files and select the product to apply product details. View the all product list.

USER BUY PRODUCT: User first registers their account and login their account. Next search the product and select to add cart. View the product details and buy the product. Give review for that selected product.

OBJECTIVE: In public cloud computing, the clients store their massive data in the remote public cloud servers. Since the stored data is outside of the control of the clients, it entails the security risks in terms of confidentiality, integrity and availability of data and service. Remote data integrity checking is a primitive which can be used to convince the cloud clients that their data are kept intact. In some special cases, the data owner may be restricted to access the public cloud server, the data owner will delegate the task of data processing and uploading to the third party, for example the proxy. On the other side, the remote data integrity checking protocol must be efficient in order to make it suitable for capacity-limited end devices. Thus, based on identity-based public cryptography and proxy public key cryptography, we will study ID-PUIC protocol.

MOTIVATION: In public cloud environment, most clients upload their data to PCS and check their remote data's integrity by Internet. When the client is an individual manager, some practical problems will happen. If the manager is suspected of being involved into the commercial fraud, he will be taken away by the police. During the period of investigation, the manager will be restricted to access the network in order to guard against collusion.

In public cloud, remote data integrity checking is an important security problem. Since the clients' massive data is outside of their control, the clients' data may be corrupted by the malicious cloud server regardless of intentionally or unintentionally.

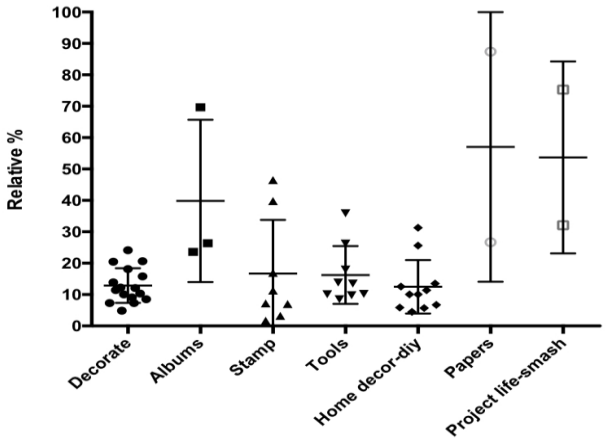
DATA PROCESSING

The initial step of web usage mining analysis is data preprocessing 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 preprocessed to discard uninteresting requests, to identify user sessions and to prepare the log to enable its analysis.

The web logs follow the Common Language Format standard (CLF) and provide raw information such as the IP address from which the session was established, the date and time of the request, the page URL or the HTTP status returned to the client, for instance. Figure 4 shows a piece of the raw web log of the considered use case, corresponding to two months of usage of the Up And Scrap website, containing 8; 607; 625 events. The preprocessing step can be split into three main phases. The first two are common to any web usage mining project The third one is introduced to prepare the log contents for applying the used model checking techniques. Let us describe that phases in more detail.

IV. IDENTIFYING USERS' BEHAVIOURAL PATTERNS

Next we are going to detail the process carried out to analyze and identify behavioral patterns from the Up&Scrap logs. Prior to the analysis we have defined a set of variables and macros based on the sets identified in Section III-D. They areenumerated below, according to their equivalence with the sets proposed



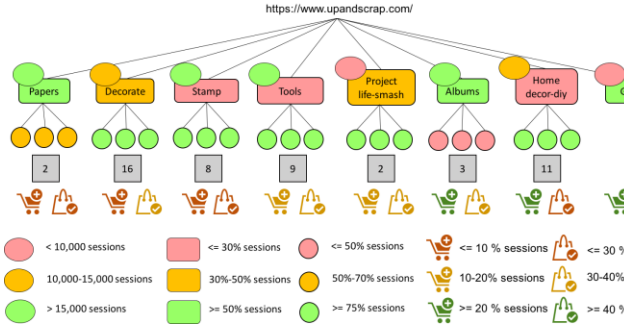
Behavioral patterns related to the buying process:
Regarding the buying process there are two specific actions that we are interested in. First, user sessions showing interest in acquiring a specific product. That corresponds to the events of adding a product to the cart and adding a product to the wish 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 to correlate such information with different access patterns. Second, sessions that buy some products, that is, sessions where the event Buy products in the cart happens.
In this regard, it is important to analyze the relation between showing interest in a product and purchasing it. First, we are interested in knowing from which sections the products are added to the cart or to the wish list.

Table Design

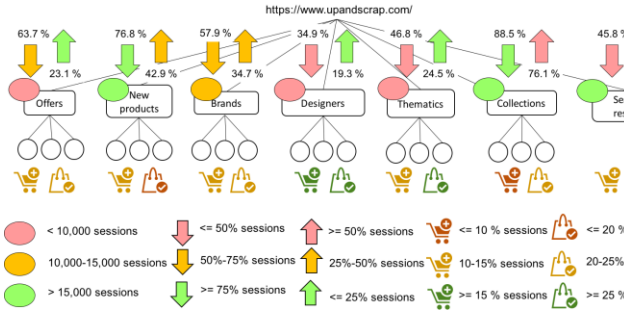
PC-PC\SQLEXPRESS...act - dbo.userreg | Summary

Column Name	Data Type	Allow Nulls
id	int	☑
username	varchar(50)	☑
password	varchar(50)	☑
mail	varchar(50)	☑
mob	varchar(50)	☑
gender	varchar(50)	☑
age	varchar(50)	☑
education	varchar(50)	☑
		☐

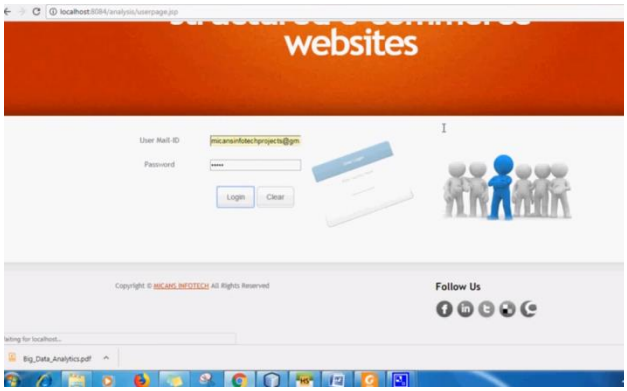
V. RESULT

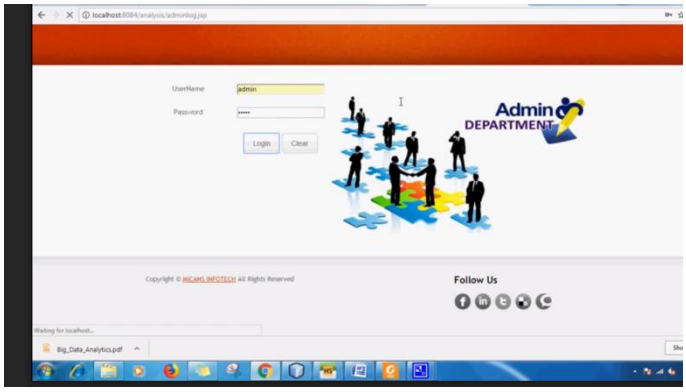


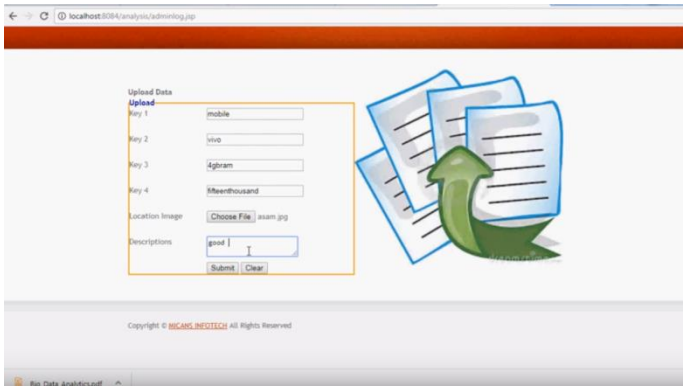
Summary of accesses to the main categories of the Up&Scrap website and their importance of secondary sections within the buying process.

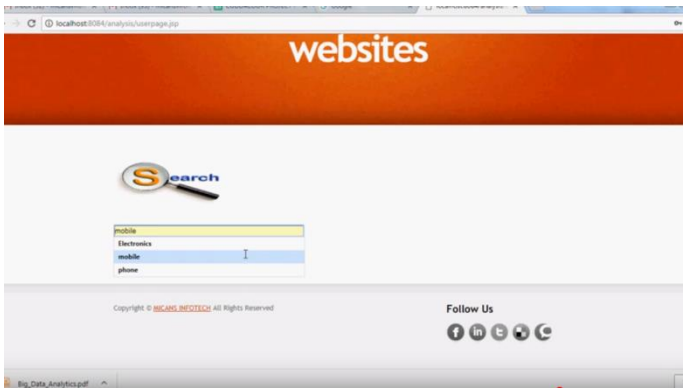


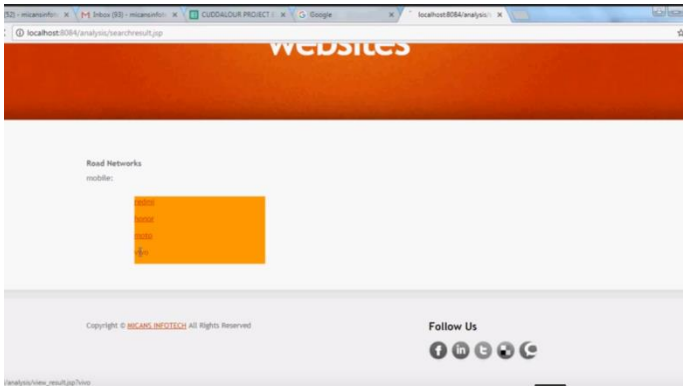
Summary of exclusive and first choice accesses to secondary sections of the Up&Scrap website and their importance of secondary sections within the buying process.

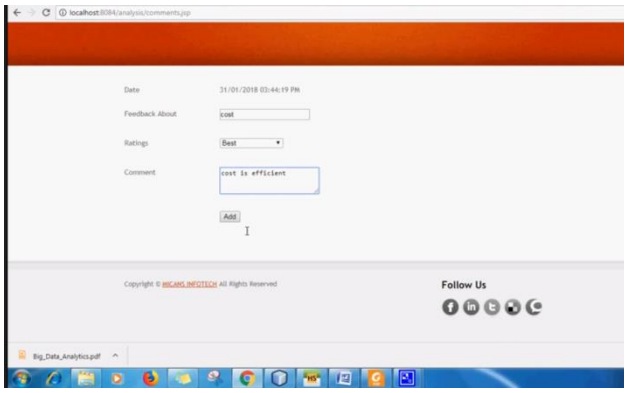


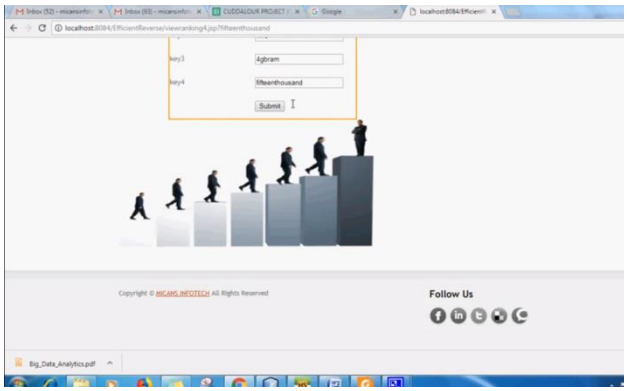


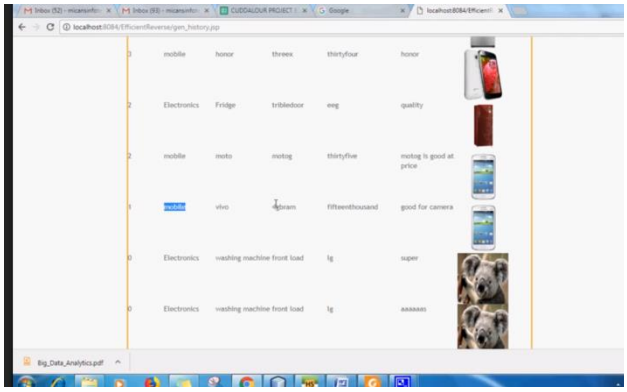












VI. CONCLUSION

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 overfitting spaghetti models or underfitting 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 analyse 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 throw 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 ecommerce 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 Knowledg 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.