LINEAR-TEMPORAL LOGIC MODEL CHECKING APPROACH FOR THE ANALYSIS OF STRUCTURED E-COMMERCE WEB LOGS

M.Joseph Anusha[1] , Venkat N. Ravipati[2]

P.G. Scholar, Dept. of CSE, Krishnaveni Engineering College for Women, Narasaraopet, Andhra Pradesh, India[1]
Associate Professor, Krishnaveni Engineering College for Women, Narasaraopet, Andhra Pradesh, India [2]

Abstract

Web based shopping is winding up increasingly basic in our every day lives. Understanding clients' interests and conduct is fundamental with the end goal to adjust internet business sites to clients' necessities. The data about clients' conduct is put away in the web server logs. The examination of such data has concentrated on applying information mining systems where a fairly static portrayal is utilized to show clients' conduct and the succession of the activities performed by them isn't normally considered. Accordingly, joining a perspective of the procedure pursued by clients amid a session can be of incredible enthusiasm to distinguish more mind boggling personal conduct standards. To address this issue, this paper proposes a direct fleeting rationale demonstrate checking approach for the investigation of organized internet business web logs. By characterizing a typical method for mapping log records as indicated by the online business structure, web logs can be effortlessly changed over into occasion logs where the conduct of clients is caught. At that point, distinctive predefined inquiries can be performed to distinguish diverse standards of conduct that consider the distinctive activities performed by a client amid a session. At long last, the handiness of the proposed methodology has been contemplated by applying it to a genuine contextual analysis of a Spanish online business site. The outcomes have recognized fascinating discoveries that have made conceivable to propose a few enhancements in the web architecture with the point of expanding its effectiveness.

Index Terms—Data mining, e-commerce, web logs analysis, behavioural patterns, model checking

I.   INTRODUCTION

In the present at any point associated world, the manner in which individuals shop has changed. Individuals are purchasing increasingly over the Internet as opposed to going conventional shopping. Web based business furnishes clients with the chance of perusing perpetual item lists, looking at costs, being ceaselessly educated, making list of things to get and getting a charge out of a superior administration dependent on their individual advantages. This expanding electronic market is exceptionally aggressive, highlighting the likelihood for a client to effectively move from one web based business when their necessities are not fulfilled [1], [2]. As a result, online business investigators require to know and comprehend buyers' conduct when those explore through the site, and also endeavoring to recognize the reasons that spurred them to

buy, or not, an item [3], [4], [5]. Getting this social information will permit web based business sites to convey a more customized administration to clients, holding clients [6] and expanding benefits [7]. In any case, finding client' conduct and the reasons that control their purchasing procedure is an extremely unpredictable assignment [3].

Web based business sites furnish clients with a wide assortment of navigational choices and activities: clients can uninhibitedly travel through various item classes, pursue numerous navigational ways to visit an explicit item, or utilize distinctive components to purchase items, for instance. More often than not, these client exercises are recorded in the web server logs [3], [8]. Web server logs store, in an arranged way, the succession of web occasions created by every client (ordinarily known as snap streams). The specific profitable clients' conduct is covered up in these logs, which must be found and broke down [9]. A right investigation can be in this manner used to enhance the site substance and structure [10], to adjust and customize substance [11], [12], [13], to suggest items [14], [15], or to comprehend the enthusiasm of clients in explicit items [16], for example.

Information mining systems have demonstrated their value for finding designs in log documents (when connected to the examination of web server logs the term web utilization mining [17] is utilized). Its fundamental objective is to find utilization designs attempting to clarify the clients' advantages. Distinctive procedures have been effectively utilized in the field of web based business, for example, order systems, bunching, affiliation rules or successive examples [18], [19]. In numerous application spaces these strategies are utilized related to process mining methods. Such strategies are a piece of the business insight space and apply explicit calculations to find shrouded examples and connections in expansive informational indexes [20]. An online business site is an open framework where any client conduct is conceivable. This adaptability makes the disclosure of a procedure arranged model speaking to clients' conduct a troublesome undertaking [21]. This is so in light of the fact that there are such a significant number of various conceivable connections that the last procedure demonstrate is either an over fitting spaghetti show or an under fitting bloom display [20], from which no helpful investigation should be possible. As an outcome, information mining strategies have been favored for the investigation of online business sites. All things considered, the present information mining methods and instruments have some obliges from the examination perspective. From one perspective, they don't work straightforwardly with the arrangements of occasions (the snap stream and every one of the information related to each snap) produced amid the client's route through the site, however with a deliberation of such succession, a sort of worldwide photo that overlooks causality relations. Such reflection depicts what occurred amid the session of a client by methods for an arrangement of abridged information, for example, the quantity of visited website pages the recurrence with which every item class was visited, or the time clients spend on a site page or classification, for example. Then again, most methods are just ready to order these deliberations or find straightforward connections among certain abnormal state occasions of intrigue

In this paper we propose the utilization of Temporal Logic and model checking methods as an option in contrast to information mining procedures. Such procedures have demonstrated their pertinence for open frameworks [22], [23], [24]. We propose here a system for utilizing it in organized online business sites. The objective is to examine the utilization of web based business sites and to find clients' mind boggling standards of conduct by methods for checking worldly rationale equations portraying such practices against the log demonstrate. Toward the starting, web server logs are preprocessed to separate the nitty gritty follows (successions of occasions of a client session). Occasions can be client or framework activities performed when a customer visits an item or item classification page, when he or she adds an item to the list of things to get, when the internet searcher is utilized, and so on. The business investigator can utilize an arrangement of (predefined) worldly rationale examples to detail questions that could assist him with discovering and comprehend the manner in which customers utilize the site. Considering the site structure and substance and in addition the diverse sorts of client's activities, these inquiries can check the presence of complex causality connections between occasions contained in the customer sessions. From the instrument perspective, the need of having control in transit the checking calculations are connected, and in addition the frustrating execution results we acquired when utilizing some model checking devices available to us, chiefly when utilized against huge models, drove us towards the enthusiasm of building up an explicit model checking device. We did it utilizing the SPOT libraries for LTL display checking [25].

As an utilization instance of the proposed methodology we portray the investigation did for the Up&Scrap1 web based business site, a vital on-line Spanish supplier of scratching items. The instance of study portrays the manner in which crude logs have been prepared, how the follows have been separated, how clients' standards of conduct have been defined and checked against the log. We additionally give some conceivable elucidations of the outcomes got for the questions and additionally some conceivable activities which could help in the re-plan of the site whose point is to enhance it.

II.   Existing System:

With the fast advancement of Internet in China, the industry's plan of action has changed. At present, incredible process has been made in Web internet business stage for its benefit and exchange quick. Rivalry for clients is the key factor for online business in the inexorably savage rivalry. On the off chance that you can get a handle on client needs, create focused on business exercises, not exclusively can give helpful exchanging mode and a wide decision for clients, yet additionally make the internet business to hold clients better. One of the arrangements is Web information mining innovation. We can get the client conduct from the perusing conduct of clients on Web and further examination, at that point to discover an answer. This will permit dealers find out about their clients' needs, and give customized by client inclinations, at that point gets the upper hand.

III.   Proposed System:

IJMTARC

In the present at any point associated world, the manner in which individuals shop has changed. Individuals are purchasing increasingly over the Internet as opposed to going conventional shopping. Web based business furnishes clients with the chance of perusing perpetual item inventories, contrasting costs, being ceaselessly educated, making list of things to get and getting a charge out of a superior administration dependent on their individual advantages. This expanding electronic market is profoundly focused, including the likelihood for a client to effortlessly move from one online business when their necessities are not fulfilled. As a result, internet business experts require to know and comprehend shoppers' conduct when those explore through the site, and also attempting to recognize the reasons that persuaded them to buy, or not, an item. Getting this social information will permit internet business sites to convey a more customized administration to clients, holding clients and expanding benefits.

➢ In the characterization contains the web browser used by the customer, the number of visited WebPages, the time the customer spent on each page, or the keywords used in search engine; focus on the users' interest in the different product categories and their characterization consist of the list of visited categories and the frequency of such visits.

➢ Unlike the previous approaches, uses text mining techniques to discover the most frequent words contained in the Web pages a customer visits, generating the session characterization from these words. This solution tries to identify the user's interests from the contents of the visited pages.

➢ Clustering algorithms are generally used to discover the sets of sessions showing a similar behavior or some common interests.

➢ This information can subsequently be used to improve the website contents and structure, to adapt and personalize contents to recommend products to understand customers' behavior related to the buying process or to understand the interest of users in specific products.

➢ Other specialists apply elective mining systems to foresee the client's conduct. Concentrate the clients' navigational successions to make measurable and probabilistic models ready to foresee the client next snap. These models are spoken to as Markov chains. By and by, these methodologies present a few downsides: the way toward making these models is computationally over the top expensive, and, moreover, this kind of models reacts to momentary thinking (the model does not have data to know how the current navigational state has been come to and how future states speaking to long haul objectives can be come to). The mix of bunching calculations and Markov chains enhances the forecasts of these factual models, as appeared in. The thought is to initially bunch client sessions applying some grouping calculations and, after, to create an explicit Markov chain for every one of the got bunches. As of now, there are ground-breaking business apparatuses for dissecting logs of web based business sites, being Google Analytics one of the primary ones. Google Analytics controls the system movement, gathers data about client sessions (first and last site page visited, pages visited, time spent on each page, and so on.), and shows reports integrating clients' conduct. These activity based information can likewise be joined with other clients' close to home and geographic

data. Google Analytics can't import the web server logs of a site, yet it works examining the data gathered by methods for page labeling procedures. Another intriguing component of the pursued mining approach is the reality of having the capacity to investigate arrangements of point by point occasions. The reality of thinking about the causal relations of occasions inside a client session, permitting searching for intra-session designs (and not just examples rehashed in various sessions) can furnish the experts with a significantly more definite point of view of a client conduct.

IV.   Module Description:

1. Clustering Module:
2. Behavioral Module:
3. linear-temporal logic model:

4.1 Clustering Module:

Clustering algorithms are generally used to discover the sets of sessions showing a similar behavior or some common interests.

4.2 Behavioral Module:

Conduct learning will permit web based business sites to convey a more customized administration to clients, holding clients and expanding benefits. The objective is to break down the use of online business sites and to find clients' mind boggling standards of conduct by methods for checking fleeting rationale recipes portraying such practices against the log mode

4.3 linear- temporal logic models:

Straight fleeting rationale or direct time worldly rationale is a modular transient rationale with modalities alluding to time. In LTL, one can encode formulae about the fate of ways, e.g., a condition will inevitably be valid, a condition will be valid until the point that another reality turns out to be valid, and so on. It is a section of the more mind boggling, which moreover permits spreading time and quantifiers. In this manner LTL is now and again called propositional fleeting rationale; abridged straight transient rationale is a section of S1S monadic second-arrange rationale of one successor

V.   Architecture Diagram:
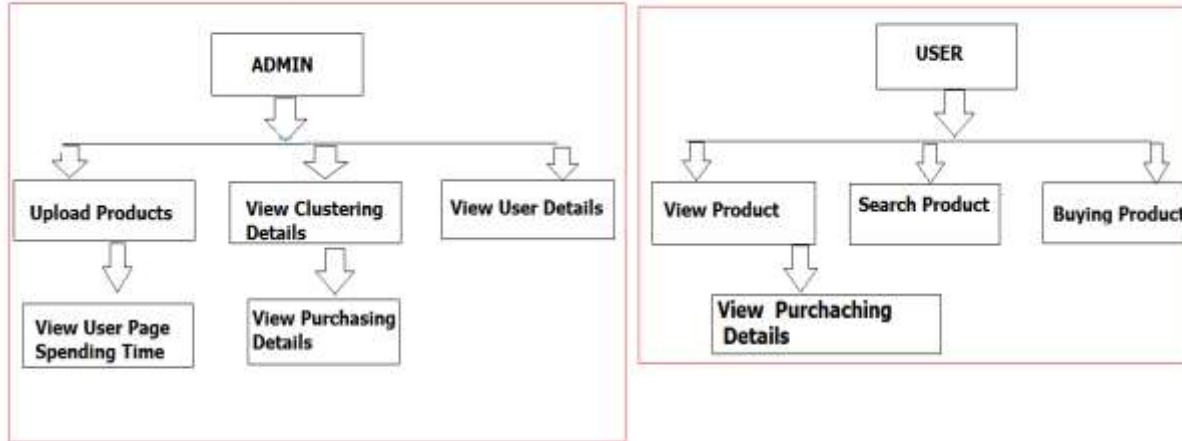
Clustering algorithm:

Bunch investigation or grouping is the assignment of collection an arrangement of articles so that objects in a similar gathering (called a group) are more comparative (in some sense or another) to one another than to those in different gatherings (bunches). It is a primary undertaking of exploratory information mining, and a typical method for measurable information examination, utilized in numerous fields, including machine learning, design acknowledgment, picture investigation, data recovery, bioinformatics, information pressure, and PC illustrations.

VI.   CONCLUSIONS

On account of open frameworks, where the groupings of collaborations (put away as framework logs) are not compelled by a work process, process mining methods whose goal is to extricate a procedure model will more often than not give either over fitting spaghetti models or under fitting blossom models, from which small intriguing data can be removed. A more adaptable methodology is required. In the paper we apply LTL-based model checking strategies to break down web based business web logs. To empower this investigation, we have proposed a typical method for speaking to occasion types and traits considering the online business web structure, the item arrangement and the conceivable outcomes of clients to explore toss the site as indicated by such association. From this basic perspective, the investigation completed has enabled us to recognize a few issues and to propose upgrades with respect to the item classification and the association of a portion of the site areas, which have been exchanged to the undertaking chiefs. In spite of the fact that the paper is emphatically identified with that site, the proposed methodology is general and the technique is relevant to organized web based business sites. The primary period of the approach, the preprocessing stage, is the one which is explicit for every online business site, since it relies upon the explicit framework log and, in the interim the examination system and the questions can be totally reused. It tends to be executed in parallel,

conveying diverse parallel servers with various parts of the log and executing the inquiries in parallel. We additionally plan to broaden the arrangement of considered examples with the end goal to break down more personal conduct standards and to encourage their programmed disclosure. For that, a one next to the other work with pros of the issue space is required with the end goal to characterize an arrangement of fascinating questions as wide as could be expected under the circumstances. Furthermore, expanding the web server logs with data about clients or online client audits will be considered. Client's data would enable us to ponder multi session examples and associate outcomes with statistic data; while, online audits would enable us to break down client's inputs with the end goal to prescribe items.

VII. REFERENCES

[1] J. B. Schafer, J. A. Konstanz, and J. Riel, "E-commerce recommendation applications." Hingham, MA, USA: Lower Academic Publishers, Jan. 2001, vol. 5, no. 1-2, pp. 115–153.

[2] N. Pogo, D. Carrere, R. Gaveled, J. Torres, and E. Agade, "Char- ´ cauterization 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. Kowhai, "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 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´ırez-Gallego, F. Torres, E. Bernal, M. J. del Jesus, and S. Garc ´ ´ıa, "Web usage mining to improve the design of an ecommerce 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 ecommerce," 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. Rebon, 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.

[15] K.-J. Kim and H. Ahn, "A recommender system using {GA} kmeans clustering in an online shopping market," Expert Systems with Applications, vol. 34, no. 2, pp. 1200 – 1209, 2008.

[16] Q. Su and L. Chen, "A method for discovering clusters of e-commerce interest patterns using click-stream data," Electronic Commerce Research and Applications, vol. 14, no. 1, pp. 1 – 13, 2015.

[17] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data," SIGKDD Explor. Newsl., vol. 1, no. 2, pp. 12–23, Jan. 2000.

[18] Q. Zhang and R. S. Segall, "Web mining: a survey of current research, techniques, and software," International Journal of Information Technology & Decision Making, vol. 7, no. 04, pp. 683–720, 2008.

[19] B. Singh and H. K. Singh, "Web data mining research: a survey," in Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on. IEEE, 2010, pp. 1–10.

[20] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes, 1st ed. Springer Publishing Company, Incorporated, 2011.

[21] N. Poggi, V. Muthusamy, D. Carrera, and R. Khalaf, "Business process mining from e-commerce web logs," in Proceedings of the 11th International Conference on Business Process Management, ser. BPM'13. Berlin, Heidelberg: Springer-Verlag, 2013, pp. 65–80.

[22] F. M. Maggi, R. P. J. C. Bose, and W. M. P. van der Aalst, Efficient Discovery of Understandable Declarative Process Models from Event Logs. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 270– 285.

[23] M. Raim, C. Di Ciccio, F. M. Maggi, M. Mecella, and J. Mendling, ¨ "Log-based understanding of business processes through temporal logic query checking," in On the Move to Meaningful Internet Systems: OTM 2014 Conferences: Confederated International Conferences: CoopIS, and ODBASE 2014, Amantea, Italy, October 27-31, 2014, Proceedings, R. Meersman, H. Panetto, T. Dillon, M. Missikoff, L. Liu, O. Pastor, A. Cuzzocrea, and T. Sellis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 75–92.

AUTHORS' BIOGRAPHY

M.Joseph Anusha received the B.Tech degree in Electronics and Communication Engineering in the year 2014 and pursuing M.Tech degree in Computer Science and Engineering from Krishnaveni Engineering College for Women.

Venkat N. Ravipati received his M.Tech degree in Computer Science and Engineering and he is currently working as Associatet Professor in Krishnaveni Engineering College for Women.