Users' Interest Analysis in E-Commerce Websites

Anup Madanshetty[1], Anil A. Bharate[2]

[1]M.Tech Student, Department of Computer Science and Engineering, BLDEA's V.P. Dr.P.G. Halakatti College of Engineering & Technology Vijayapur, Karnataka, India. E-mail: anup.mshetty16@gmail.com

[2]Assistant Professor, Department of Computer Science and Engineering, BLDEA's V.P. Dr.P.G. Halakatti College of Engineering & Technology Vijayapur, Karnataka, India. E-mail: anil.bharate@bldeacet.ac.in

ABSTRACT

Web based shopping is ending up increasingly normal in our day by day lives. Understanding clients' interests and conduct is basic so as to adjust web based business sites to clients' prerequisites. The data about clients' interest is put away in server log. Investigating such a data has concentrated over application of information mining procedure which includes fairly static portrayal to demonstrate clients' conduct & arrangement of activities performed by them isn't considered typically. Accordingly, joining the perspective of process took after by the clients amid a session can be of extraordinary intrigue in distinguishing unpredictable behavioral examples. As a solution to the problem, here we propose an approach of linear temporal logic to investigate organized web logs of online business. Thereby characterizing the typical method for mapping log records concurring to web based business structure, changing over web logs into event logs from which conduct of the clients will be caught. At that point, diverse predefined questions can be performed to distinguish distinctive behavioral examples that think about the diverse activities performed by a client amid a session. Thus the outcomes have recognized fascinating discoveries that have made conceivable to propose a few upgrades in the web composition with point of expanding its productivity.

INTRODUCTION

In the present associated world the style of individuals shopping has changed. Individuals are purchasing increasingly over Internet as opposed to going customary shopping. Web based business gives clients the chance of perusing interminable item inventories, contrasting costs, being constantly educated, making list of things to get charge out of a superior administration in view of the individual advantages. The electronic market thus expanded is much focused, highlighting likelihood of a client to effortlessly make a move from web based business when they are not fulfilled with their necessities. As an outcome, internet business experts require to know and comprehend shoppers' conduct when they explore through the site, and also endeavoring to recognize the reasons that persuaded them to buy, or not, an item. Getting this behavioral learning will permit internet business sites to convey a more customized administration to clients, holding clients and expanding benefits.

RELATED WORK

In the web based business field, most of the information mining procedures process server logs to separate sequence of the clients' navigation event. Whereas those events will not be straightforwardly undergo mining; rather, every such event changed as classification of a session. This typically comprises a set of high level state information summary that occurs amid a client's navigation. Structures of such corresponding contents may differ. While classification in [10] contains the information about number of pages visited, duration for which client was on every page he visited or keywords used in search engine. Whereas [14], [15] center around clients interest for diverse item classification that comprise of list of categories visited & the recurrence of such visits. Dissimilar to the prior techniques, [16] utilizes text mining techniques for finding most frequently used words over a web page visited by the client, producing session classification of such words. This method tries to derive the clients' interest based on contents of pages visited. Another method classifies clients based on answer provided to some set of questions [17] or utilizes a blend of clients buying & individual information. Regardless, once client classification has been figured, clustering algorithms are for the most part utilized to derive a set of sessions demonstrating a comparable conduct or some normal interests.

EXISTING SYSTEM

Presently, for convenience and faster transactions in e-commerce platform much advancement has been made. Attracting users plays a very important role in e-commerce business. If one can get to know customer interests then business activities can be developed accordingly, which will provide excellent trade mode, wide range of choices for clients but also helps in retaining them. Hence web data mining technology provides one of the best solutions.

Current systems usually derive interest of client interest through browsing habits over the web and further processes will let sellers know about their clients' needs, for which they can provide some recommendations, which will finally lead to competitive advantage.

PROPOSED SYSTEM

In the present at any point associated world, the way individuals shop has changed. Individuals are purchasing increasingly finished the Internet as opposed to going conventional shopping. Web based business furnishes clients with the chance of perusing unending item lists, looking at costs, being persistently educated, making list of things to get and getting a charge out of a superior administration in light of their individual advantages. The expanding market of electronics is profoundly aggressive, including likelihood to effectively transit from current to various other web based business whenever they feel they are not satisfied. Which leads, web based business experts to know & comprehend buyers' interest whenever they explore the site & in addition endeavoring to distinguish the reasons that spurred them to buy, or not, an item. Getting this behavioral learning will permit online business sites to convey a more customized administration to clients, holding clients and expanding benefits.

ANALYSING EVENT LOGS USING MODEL CHECKING

A. Linear Temporal Logic (LTL) & Model checking fundamentals

Discussing conduct of program requires having the capacity to discuss states of the program & furthermore advancements in states. LTL has been utilized for discussing these advancements. We utilize temporal logics to discuss states $((v_1 \wedge v_2) \vee \neg v_3)$. The execution of program may be viewed as a sequence of ordered grouping of Boolean equations fulfilled by such progressive states reached by program. Such an arrangement of execution is taken as temporal structure. That has limited arrangement of conceivable executions of program permits investigation of program behavior. Model checking procedure has been introduced to do such examination. This strategy checks truth of an arrangement of behavioral details, expressed as temporal equations, against model of system, that is made out of arrangement of conceivable executions [18][19]. Thus we will utilize LTL that characterizes logic for (infinite) program execution traces with the approach followed in [20].

B. Event log analysis by application of model checking

We now shift from conceptual framework to instance of event log examination. The traces are supposed to be keep running of program, here arrangement of atomic propositions compares with arrangement of event properties or events. Each line in event log corresponds to event & columns relate to attributes of event. Events are arranged by the time each one occurred. We are thinking about the arrangement of events relating to a similar session, those with a similar Id. We relate a atomic proposition to every value of attribute showing up in events, the entire set is said to be α. The event sequence for trace might be viewed as the elements sequence associated to 2^α. Keeping in mind the end goal to empower the utilization of model checker method, traces that are finite, have to be changed to infinite ones. To accomplish this, there are diverse recommendations in literature. The most regularly utilized is the addition of a dummy End event to each terminal state. By this we actually, for every trace dummy final state a transition has been added, and also a self-iterating loop of state, both named after, conjunction of all the Atomic variables & End event are negated. The model checker should be careful about order of transformation for avoiding errors in interpretation.

C. Application of the display checker

Keeping in mind the end goal to empower the utilization of LTL model checking over event logs, we built up a log analysis framework made out with two principle segments are offered as REST web services. Deeper insights of model checker are found in [21].

     1] Model Generator loads and transforms input log record, indicated as a Comma Separated Values (CSV) document, in order to feed into checker.

     2] The Model Checker, that analyses & loads previous document. The model checker is developed by utilizing the SPOT libraries.

Other than using LTL formulas, this tool provides likelihood of characterizing sets of macros & variables for the making of less demanding LTL equations. Variable subsets are characterized by numerous ways: enumerating, as the scope of identifier or using regular expressions. When VAR set is defined, presence of "?VAR" in the formula

implies that this equation should be assessed for every component associated with "VAR", so that, the same number of equations as components in VAR set are naturally checked by the tool.

D. Applying model checking to analyze the online business sites

Clients of any internet business webpage explore through the diverse pages by execution of two types communications: either a GET activity to recover some data or a POST task, ordinarily asking for the site to execute some activity, for example, adding some item to the cart, getting some item, signing in, and so on. The site log records such activities together with some related data, for example, the IP address the client is associated to or the time at which the communication happened, for example.
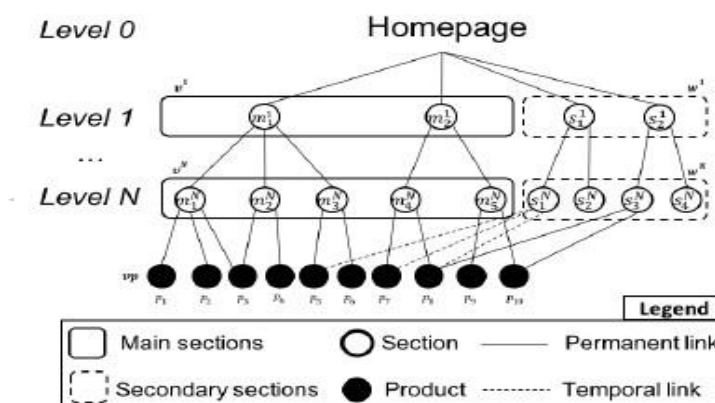

Figure 2: Usual structure utilized for product categorization in e-commerce site.

To apply demonstrate checking systems, we will relate LTL equations to events, that enables us to view log as a Kripke structure speaking to the model to be examined. Hence, we will be characterizing the set of Atomic propositions, and change events into conjunctions of these variables. Which will be finished amid the pre-preparing stage, whose yield is model representing the log. Fig 2 demonstrates average structure utilized as a part of internet business sites to compose and sort items. Comparable scientific categorizations have been proposed by various taxonomies yet including just fundamental areas. From the home page (level 0) distinct segments can be accessed (level 1). Two distinct sorts of segments may be recognized.

1) Main section that compares the principle item order. Such segments enable in accessing every item. All in all, the item order is typically disjoint, yet it isn't obligatory: in some online business sites a similar item could have a place with various sections.

2) Secondary section, which give an optional arrangement of the site items, whose goal is to enable the entrance to a subset of items with some normal and particular highlights. Not at all like the past case, must not every item be available from those sections. Moreover, recognizing two distinct kinds of secondary sections relying upon whether items within those segments are for all time or incidentally placed in the section. A case of segments with brief items would be offers or segments with new items that are intermittently reestablished. A case of optional areas with changeless links having sections where user gets to items by producer, theme, and so on.

Characterizing diverse regular occasions and sets of occasions empower the likelihood of proposing design inquiries that can be investigated autonomously of the web based business site dissected. As expressed above, demonstrate checking can be utilized to question about particular expresses, their development and their relations. With regards to online business web server logs this implies demonstrate checking may be utilized to break down areas visited by clients, navigation ways took after while getting to particular sites pages, the connection of various web segments or segments that lead clients to purchase items, as an example. In light of the characterized structure and sets, we will characterize three unique kinds of question designs.

1) Queries identified with the investigation of the web pages went by clients. This first sort contains the most straightforward inquiries: which segments do the clients visit? Cases of fascinating questions in such manner would be: Sessions where the client visits fundamental level-1

- Sessions, for any $v \in V$ [1], the LTL equation (v) will give the quantity of follows in which the client visits classification e of the primary segments of level 1.

- Sessions where the client visits a portion of the optional areas: $(w^1 \lor w^2 \lor \ldots \lor w^N)$ will tally what number of sessions contain no less than one occasion with one of the auxiliary atomic propositions.

- Sessions where clients only visit level-1 principle segments (no optional segment is gone to): $(v^1) \land \neg (w^1 \lor w^2 \lor \ldots \lor w^N)$

2) Queries distinguishing navigational examples. These questions attempt to build up causal relations between he realities of going by various segments in the website:

- specific segments visited immediately after level-1 main section: for any $e \in (V \cup W)$, $(v^1 \land 0^e)$ include the quantity of traces which the next event in the wake of going by primary level-1 segment contains e.

3) Queries corresponding to client navigational examples with client activities. These questions endeavor to correspond particular activities, unique in relation to going to segments, with the behavioral examples beforehand recognized. Case of these questions is the accompanying ones:

- Sessions that play out a particular activity without going by primary areas: for any $a \in A$: $\neg (v1 \lor v^2 \lor \ldots \lor v^N)$ (a) checks the quantity of sessions where the activity is performed without going to any main section.

CONCLUSION

On account of open frameworks, where interaction sequence are not workflow constrained, the goal of process mining strategies is the extraction of a process model, will more often than not gives either a over fitting spaghetti models or under fitting flower models, that will lead to extraction of very small amount of interesting data. Therefore a more adaptable approach is needed therefore we use LTL based model checking procedures over weblogs of e-commerce websites to analyze them. To empower this analysis, we propose a typical method for representing types of events and its attributes. Considering the structure of website, the product classification and the potential outcomes of clients to explore the site as per such association. The approach proposed is generalized technique that is relevant web based business sites. The primary phase of method, the preprocessing stage, is the one which is particular for every online business site, since it relies upon system logs that are specific &, therefore analysis procedure & queries are totally reusable. Queries will be parallel executed, by sending diverse parallel servers with various blocks of log file and parallel query execution. Also we plan to extend studied patterns sets so that to examine more behavioral & automatic discovery facilitation. Moreover, expanding the web server logs with data about clients or online client audits will be contemplated. Client's data would enable us to think about multi session patterns and correspond comes about with demographic data; while, online audits would enable us to break down client's criticisms so as to suggest items.

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 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 Jes´us, 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] 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

[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.

[15] 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.

[16] . 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.

[17] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach,"Electronic Commerce Research and Applications, vol. 10, no. 3, pp. 331–341, May2011.

[18] E. M. Clarke, A. Emerson, and A. P. Sistla, "Automatic Verification of Finite State Concurrent System Using Temporal Logic Specifications: a Practical Approach," in Proceedings of the 10th ACM SIGACT-SIGPLAN symposium on Principles of programming languages (POPL 1983), Austin, Texas, Jan 24 - 26, 1983. New York, NY, USA: ACM, 1983, pp. 117- 126.

[19] E. Clarke, O. Grumberg, and D. Long, "Verification tools for finite-stateconcurrent systems," in Workshop/School/Symposium of the REX Project (Research and Education in Concurrent Systems). Springer, 1993, pp.124175.

[20] J. Couvreur, "On-the-fly verification of linear temporal logic," in Proceedings of Formal Methods: World Congress on Formal Methods in

the Development of Computing Systems, Toulouse (France), September,1999, pp. 253–271..

[21] P. A´ lvarez, J. Fabra, S. Herna´ndez, and J. Ezpeleta, "Alignment of teacher's plan and students' use of lms resources. analysis of moodle logs," in 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET), Sept 2016, pp.1–8.

[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. R¨aim, C. Di Ciccio, F. M. Maggi, M. Mecella, and J. Mendling, "Log-based understanding of business processes through temporal logic.