IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Special Issue-15, May 2019 E-ISSN: 2347-2693

Tracking The User's Behaviour in E- Commerce Website

Smriti Gupta[1*], Komal Kumari[2], Latha A[3]

[1,2,3]Department of CSE, SJB Institute of Technology, India

DOI: https://doi.org/10.26438/ijcse/v7si15.257260 | Available online at: www.ijcseonline.org

***Abstract*—** Online shopping is becoming more and more common in our daily lives. Tracking user's interests and behaviour is essential in order to fulfil customer's requirements. The information about user's behaviour is stored in the web server logs. Absorbing a view of the process followed by user's during a session can be of great interest to identify the behavioural patterns. The analysis of such information has focused on applying data mining techniques. Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. It is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. To address this issue, in this work we proposes a linear temporal logic model checking method for the analysis of structured e-commerce web logs.

Keywords: Data mining, e-commerce, web logs analysis, behavioural patterns, model checking.

I. INTRODUCTION

Data mining is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. It is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. In today's world, the way of shopping has changed. People are buying more and more over the Internet instead of going traditional shopping. E-commerce provides customers with the facilities of browsing endless product catalogues, comparing prices, creating wish list and 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. Ecommerce business analysts require to know and understand consumer's behaviour when those go through the website, as well as trying to identify the reasons that motivated them to purchase. Getting this behavioural knowledge will allow e-commerce websites to deliver a more personalized service to customers and increasing benefits. Usually, the user's activities are recorded in the web server logs. Web server logs are stored in an ordered way, the sequence of web events generated by each user is stored as a separate log. The very valuable user's behaviour is hidden in these logs, which must be discovered and analysed. In this paper we propose the use of Temporal enjoying a better service based on their individual interests. E-commerce websites Logic and model checking techniques as an alternative to data mining techniques. 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.The focus is on the user's interest in the different product categories and their characterization consist of the list of visited categories and the frequency of such visits. The goal is to analyse the usage of ecommerce websites and to discover customer's complex behavioural patterns by means of checking temporal logic formulae describing such behaviours against the log model. The business analyst can use a set of temporal logic patterns to formulate queries that could help him to discover and understand the way clients use the website. Customers reviews are considered for the further improvements.

II. EXISTING SYSTEM

With the rapid development of Internet in China, the industry's business model has changed. At present, great process has been made in Web e-commerce platform for its convenience and transaction fast. Competition for users is the key factor for e-commerce business in the increasingly fierce competition. If you can grasp customer needs, develop targeted business activities, not only can provide convenient trading mode and a wide choice for customers, but also make the e-commerce business to retain customers better. One of the solutions is Web data mining technology. We can get the user behaviour from the browsing behaviour of customers on Web and further analysis, then to find a solution. This will allow sellers know more about their customers needs, and provide personalized according to customer preferences, then obtains the competitive advantage.

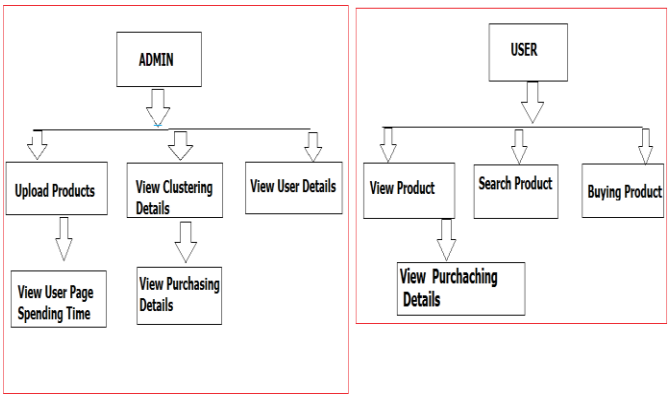

Fig. 1

III. DESIGN

ADMIN: Admin has to login first into website. Admin has privileged to see users details , can upload pictures, can view clustering.

Upload Product Picture: Admin can upload picture of the product with all the details for the user.

View Clustering: Admin has privileged to see all details of user and can see the product in user's wishlist. Admin also has privileged to see the buying details of user.

View Users Details: Admin has privileged to see all the details of users.

View Users Page Spending Time: Admin can see how much time a user is spending on a particular page.

View Purchasing Details: Admin can view the purchasing details of the user. **USER :** User has to login into website to see all the products with details and for buying the product.

View Product: Users can view all the product details and can wishlist it for the future purchase.

Search Product: User can search for any product by its name or the link.

Buying Product: User can buy the interested product by seeing its all details and other customers feedback.

View Purchasing Details: Users can see the purchasing details of the product which is buyed by the users.

IV. METHODOLOGY

Clustering Module: Clustering algorithms[1] are generally used to discover the sets of sessions showing a similar behaviour or some common interests.

Behavioural Module: Behavioural knowledge[2] will allow e-commerce websites to deliver a more personalized service to customers, retaining customers and increasing benefits.The goal is to analyse the usage of e- commerce websites and to discover customers' complex behavioural patterns by means of GET and POST methods and checking temporal logic formulas describing such behaviours against the log mode.

linear- temporal logic model:lineartemporal logic or linear-time temporal logicis a modal temporal logic[3] with modalities referring to time. In LTL, one can encode formulae about the future of paths, e.g., a condition will eventually be true, a condition will be true until another fact becomes true, etc. It is a fragment of the more complex, which additionally allows branching time and quantifiers. Subsequently LTL is sometimes called propositional temporal logic, abbreviated Linear temporal logic is a fragment of S1S monadic second-order logic of one successor.

Clustering algorithm: Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a main task of exploratory data mining, and a common technique for statistical data analysis,used in many fields, including machine learning, pattern recognition, imageanalysis, information retrieval, bioinformatics, datacompression,and c omputer graphics. The combination of clustering algorithms and Markov chains improves the predictions of these statistical models, as shown in [4].

Model Checking: Users of any e-commerce site navigate through the different web pages executing two types of interactions: either a GET operation to retrieve some information or a POST operation, usually requesting the website to execute some action, such as adding some product to the cart, buying some product, logging in, etc. The website log records such actions together with some associated information, such as the IP the user is connected from or the time at which the interaction occur [5], for instance. Some of these actions correspond to events that are common to any e-commerce website such as the ones related to visiting the sections containing products. Therefore, a general way of classifying the events in the web logs according to the product categorization can be proposed. From now on, we are going to describe the proposed approach to relate the website structure and the events in the log, to identify meaningful set of events, and to ask for behavioural usage patterns using model checking based on the previous classification.

V. RESULTS

SL NO	TITLE	ADVANTAGE	DISADVANTAGE
1.	Repeat Buyer Prediction for ECommerce	This paper present winning solution which consists of comprehensive feature engineering and model training to identify who can be converted to regular loyal buyers and then target them to reduce promotion cost and increase the return on investment.	Although the paper is strongly related to that website, the proposed approach is general and the methodology is less applicable to structured ecommerce websites.
2.	A method for discovering clusters of ecommerce interest patterns using clickstream data	The browsing behavior of a number of consumers - including frequency and time spent on each category.	Deploying different parallel servers with different parts of the log and executing the queries in parallel.
3.	Purchase influence mining: Identifying top k items attracting purchase of target item .	Based on the item graph, identifying items that attract users to purchase the target item could be used for supporting business decisions. Purchase Influence Mining finds the top k items maximizing the estimated purchase influence from them to a target item.	Structural point of view, the paper proposes a set of query patterns, translated into LTL formulas, which are of interest for the domain of electronic commerce.
4.	Predicting user behavior through sessions using the web log mining	It is the method to extract the user sessions from the given log files. each user is identified according to his/her IP address specified in the log file and corresponding user sessions are extracted	Only few percentage of traces has been done and it will not satisfy the corresponding formula, to extract interest in correlations among sequences of events.
5.	Mining customer requirements from online reviews.	In this paper, first five categories include the features to predict reviews helpfulness from the perspective of a product designer and then present an approach based on conjoint analysis to measure customer requirement.	Extend set of studied patterns is in more complex behavioural forms and it is difficulty to analyse their automatic discovery.

VI. PROPOSEDSYSTEM

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 wishlist 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' behaviour 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 behavioural knowledge will allow e-commerce websites to deliver a more

personalized service to customers, retaining customers and increasing 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 behaviour 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' behaviour related to the buying process or to understand the interest of users in specific products.
- Another researchers apply alternative mining techniques to predict the user's behaviour. extract the users' navigational sequences to create statistical and probabilistic models able to predict the user next click. These models are represented as Markov chains. Nevertheless, these approaches present some drawbacks: the process of creating these models is computationally very expensive, and, besides, this type of models responds to very short-term reasoning (the model does not have information to know how the current navigational state has been reached and how future states representing long term goals can be reached). The combination of clustering algorithms and Markov chains improves the predictions of these statistical models, as shown in. The idea is to first group user sessions applying some clustering algorithms and, after, to generate a specific Markov chain for each of the obtained clusters. Currently, there are powerful commercial tools for analysing logs of e-commerce websites, being Google Analytics one of the main ones. Google Analytics controls the network traffic, collects information about user sessions (first and last web page visited, pages visited, time spent on each page, etc.), and displays reports synthesizing users' behaviour. These traffic-based data can also be combined with other users' personal and geographic information. Google Analytics is not able to import the web server logs of a website, but it works analysing the information collected by means of page tagging techniques. Another interesting feature of the followed mining approach is the fact of being able to analyse sequences of detailed events. The fact of considering the causal relations of events inside a user session, allowing to look for intra- session patterns (and not only patterns repeated in different sessions) can provide the analysts with a much more detailed perspective of a user behaviour.

VII. CONCLUSION

In the case of open systems, where the sequences of interactions 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 under fitting flower models, from which little interesting information can be extracted. In the paper we apply LTL-based model checking techniques to analyse ecommerce web logs. We plan to extend the set of studied patterns in order to analyse more behavioural patterns and to facilitate their automatic discovery. Additionally, extending the web server logs with information about users or online customer reviews is going to be studied. User's information would allow us to study multi session patterns and correlate results with demographic information; while, online reviews would allow us to analyse customer's feedbacks in order to recommend products.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] S.D. Bernhard, C.K. Leung, V.J.Reimer, and J.Westlake, "Clickstream prediction using sequential stream mining techniques with markov chains," in Proceedings of the 20thInternational Database Engineering & Applications Symposium, ser. IDEAS '16. New York, NY, USA: ACM, 2016, pp. 24–33.

