INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)
Available online at www.ijariit.com

Big Data Classification of Users Navigation and Behavior Using Web Server Logs

Prajakta Ghavare
Department of Computer Engineering
Dr. D.Y. Patil College of Engineering, Pune, Maharashtra
prajaktaghavare@hotmail.com

Prashant Ahire
Department of Computer Engineering
Dr. D.Y. Patil College of Engineering, Pune, Maharashtra
ksprashantahire@gmail.com

Abstract: ***Users for online shopping are increasing day by day because of easy to get and time-saving property of online shopping. Having a proper understanding of users interest for certain type of product or different products for online shopping becomes important to create personalized service for a target market. An important property of successful e-commerce website is the ability to provide useful content at the right time to users. And because of all this, personalization techniques are introduced to create adaptive shopping application in which user interfaces change according to users interest. User's behavior information is stored in web log files, and to get the information data mining techniques are used in which they use statistical characters to model user's behavior and not considering the sequence of action performed by uses. It becomes helpful if we follow user's session to understand complex user behavior. Therefore to eliminates all these issues this paper proposes a linear-temporal logic model checking approach for the analysis of structured e-commerce weblogs. If we consider a common way of mapping log records according to e-commerce structure, weblogs can be easily converted into event logs by which behavior of the user is captured. After getting users behavior by performing different predefine queries to identify different behavioral patterns that consider the different actions performed by a user during a session.***

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 traditional 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 behavior 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 behavior 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 behavior 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 behavior to enhance 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 find users behavior and to find users interest. Numbers of algorithms are proposed in recent years for data mining in the field of e-commerce 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 user 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' behavior on e-commerce site to discover customers' complex behavioral patterns by means of checking temporal logic formulas describing such behaviors against the log model. At the start using web server log user behavior is generated. After generation business analyst can use set of predefined queries which help him to discover the way client use the website.

II. BACKGROUND

Many different approaches have been proposed till now to discover user's behavior for a recommendation or to analyze the site for a better approach to the user. [3] Includes such a method for e-commerce on the basis of clickstream. In which they discover users behavior on the basis of a number of customers their browsing sequence, the frequency of item visited by the customer, how much time customer spend on each category etc. [4] is work on process mining or process discovery. Process mining technique automatically generates a complex business project depending on examples execution on event logs. Data mining technique is the data-driven technique which discovers a pattern in raw data. Web Usage Mining is one of the data mining techniques which identifies usage pattern of the web data so as to provide better web application [6]. Another data mining technique is given in [5] in which they discover that user's behavior can be studied by analyzing the log file.

III. LITERATURE SURVEY

Guimei Liu, Tam T. Nguyen et.al, proposed a method called Repeat Buyer Prediction for E-commerce. In this approach, they have generated different profiles for users, merchants, brand, categories item and their interactions through feature engineering. To generate all these features they trained different classification model which mainly contains GBM, Logistic Regression, XGBoost, Factorization Machine and Random Forest. They have also used ensemble technique to mix the different classifier to improve the performance [1].

The recommendation system is used in e-commerce to give a proper prediction to the user for buying a product. Recommendation system uses product knowledge learned from user's behavior to guide or discover a particular product which he will buy. In [2] J Ben Schafer et.al, present a study on, how recommendation system are related to conventional database analysis technique. They also develop one recommendation system which takes the input from users and from databases to give a recommendation to the user.

A method for finding out a bunch of e-commerce interest prototype by means of click-stream information is given in [3] by Qiang Su and Lu Chen. Analysis of the customer's behavior is a very important task for the business analyst to improve the application as per target market. In this for data analysis they use different browsing behavior such as a number of users, their visiting sequence, time and frequency spend by the user on each category etc. and depending on this they developed an improved clustering technique and improve it with the set theory to generate user interest pattern.

[4] Gives, Online Discovery of Declarative Process Models from Event Streams. Most of the business process is controlled and the process by the information system. Such kind of system record real-time information about business processing during execution. In this, they proposed a framework for the generation of LTL-based declarative process models from streaming event data. This framework continuously updates a set of valid business constraints based on the events occurred in the event stream. A declarative model is presented with the help of Declare, which combines a formal schematic features Linear Temporal Logic (LTL) on finite traces, with a graphical notation.

Leandro G. Vasconcelos et.al proposed a new e-commerce personalized technique called Exploiting client logs to support the construction of adaptive e-commerce applications. In this work, they first present that it is possible to construct personalization by observing or analyzing the customer's behavior while browsing e-commerce web site. For this hypothesis, they built one application which allows automatic generation and analysis of log file on the real-time basis [5].

Generally, there is two type of log files are generated one is a server sided log file and other is a client-side log file. The server-side log file is automatically generated by server and client side log file is managed accurately for user analysis. This include three stages first is data cleaning, second is user identification and last is session identification. Depending on these three stages G. Neelima et.al developed one application using the Weblog mining called Predicting user behavior through Sessions using the Weblog mining. In this method, they extract user's information from given log files. At the start, each user is recognized by his or her IP address and from that their session is generated. After extracting session frequency of user visited the particular page is extracted and from this, they analyze the user's behavior [6].

[7] Gives a method for e-commerce data mining called Web usage mining to improve the design of an e-commerce website. In this, they present some set of stages which includes data collection, data processing, extraction of useful data and its analysis etc. The useful data is extracted by supervised and unsupervised data mining algorithms with the help of some task such as clustering,

association and subgroup discovery. The result of all this process is then discussed with designing team to improve the e-commerce website.

Most of the studies till now for the product recommendation only consider past purchased data i.e historical data of the users. However, some method considered users navigational and behavioral data for the recommendation. On the basis of this concept [8], Yong Soo Kim and Bong-Jin Yum proposed a new approach called Recommender system based on clickstream data using association rule mining. In this novel approach, they improve the proposed collaborative filtering method by calculating the confidence level between clicked products, between the products placed in the basket, and between purchased products, respectively. After calculation confidence level, preference level is calculated by using the above confidence level calculation.

Roung-Shiunn Wu and Po-Hsuan Chou [9] proposed an approach named Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. For the e-commerce websites, it is necessary to analyze the user's behavior. Online customer's segmentation is the process of dividing customers into multiple categories which contribute to better characterization and understanding. Therefore to get the proper segmentation author developed a soft clustering method which uses hidden mixed-class membership clustering approach to differentiate users on the basis of their purchasing data across categories. For the segmentation, they used hidden Dirichlet allocation model.

K. Sudheer Reddy, M. Kantha Reddy el.al proposed a method for web usage mining. Web Usage Mining is one of the categories of data mining which identifies usage pattern to perceive and better serve the requirements of the web applications. The web usage mining consists of three stages i.e processing, pattern discovery, and analysis. Data processing is the main and essential process in web usage mining which helps to improve the quality of data mining. In this study, they present different data presentation method to access streams prior to the start of the mining process to improve the data processing for unique identification of user and session [10].

IV. CONCLUSION

In this paper, a study based on a linear-temporal logic model checking approach for the analysis of structured e-commerce web logs are presented. To analyze the user's behavior the proposed method represents event type and attribute according to e-commerce web structure. And from this structure, a set of query set is proposed. From these set of queries correlation among the sequence of events is found to analyze the user behavior on e-commerce web site.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] A. Burattin, M. Cimitile, F. M. Maggi, and A. Sperduti, "Online discovery of declarative process models from event streams," IEEE Transactions on Services Computing, vol. 8, no. 6, pp. 833–846, 2015.

[5] L. G. Vasconcelos, R. D. C. Santos, and L. A. Baldochi, "Exploiting client logs to support the construction of adaptive e-commerce applications," in 2016 IEEE 13th International Conference on e-Business Engineering (ICEBE), 2016, pp. 164–169.

[6] G. Neelima and S. Rodda, "Predicting user behavior through sessions using the web log mining," in 2016 International Conference on Advances in Human Machine Interaction (HMI), 2016, pp. 1–5.

[7] 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.

[8] 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.

[9] 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, May 2011.

[10] K. S. Reddy, M. K. Reddy, and V. Sitaramulu, "An effective data preprocessing method for web usage mining," in Information Communication and Embedded Systems (ICICES), 2013 International Conference on. IEEE, 2013, pp. 7–10.

